AN APPLICATION OF PARETO DISTRIBUTION TO THE STUDY OF THE STRUCTURE OF WAGES AND SALARIES IN SOME SELECTED MANUFACTURING INDUSTRIES

A THESIS

Presented to

The Faculty of the Graduate Division

by

Apel Dido

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Industrial Engineering

Georgia Institute of Technology

May, 1963

AN APPLICATION OF PARETO DISTRIBUTION TO

THE STUDY OF THE STRUCTURE OF WAGES AND SALARIES

IN SOME SELECTED MANUFACTURING INDUSTRIES

Approved:

Date approved by Chairman: May 30, 1963

ACKNOWLEDGMENTS

The author expresses his sincere appreciation to those who have helped make this study possible, especially to Dr. Joseph Krol for his interest, direction, and assistance in completing this thesis. Many thanks are extended to Dr. Harrison M. Wadsworth and Dr. William C. Biven, members of the reading committee, for their advice and participation in this research. For the cooperation of the companies, without whose participation this work could not have been done, the author is extremely grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (CONTINUED)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (CONTINUED)

SUMMARY

The primary objective of this study was to ascertain whether the Pareto distribution could be usefully applied as a quantitative method of comparison in analyzing the structure of wages and salaries in some selected manufacturing industries.

The secondary objective of this study was to investigate the distribution of wages and salaries among various manufacturing industries and individual companies within these industries; and further, to illustrate the econometric methods of approach as a useful tool to the industrial engineer.

This survey of employee earnings represents companies having 100 or more paid employees, and is limited to participating firms from 14 manufacturing industry groups. Statistical income data were gathered by confidential questionnaires mailed to the wage and salary administrators of 211 companies. Replies were received from 19 companies; due to being incomplete, four of these were excluded from the study.

Employing the method of least-squares, income data of each of the 15 companies were fitted to the Pareto distribution expressed by the formula $y = a\,x^{-v}$. In this formula, y represents the number of persons receiving an income of x or more, and a and v represent constant parameters. In obtaining the desired curve for a given industry, data of all the establishments within the same industry were combined. To illustrate the closeness of fit between the observed values of y and the

computed values of y, these values were plotted on full logarithmic paper.

Appropriate methods of curve fitting were explored to obtain a simple formula, which may or may not have a physical basis, but by which an approximate value of y may be computed from a given value of x. Chapter V discusses three methods of attack undertaken to describe the entire range of incomes by empirical means.

The results obtained were favorable as they indicated that the Pareto distribution generally fits well limited portions of the observed income curve, with different values of v. The upper salary range of the income curve for a given company can be adequately represented using the Pareto distribution. The parameters a and v portray the income distribution characteristics of the company. Based on these parameters, the structure of salaries can be analyzed comparatively in the manufacturing industries among various companies.

It is recommended that income data be collected from many companies accurately and in more complete form; then analyses of similar nature be undertaken. Future studies of this type should attempt to ascertain the existence of a "pattern" in the distribution of incomes with the hope that some day the income distribution patterns and their cause-and-effect relationships will be understood quantitatively.

CHAPTER I

INTRODUCTION

The subject of economic analysis has been divided into two parts, namely, economic statics and economic dynamics. Economic statics deals with the problems of equilibrium conditions, which do not involve variations in the time variable. However, economic phenomena which depend upon their sequence in time, as in the theory of business cycles, belong to economic dynamics.

So far as this study is concerned, the reader is requested to accept the view that those problems which are concerned with equilibrium in the economy, that is to say, problems which may be described in terms of mathematical models that do not introduce the variable of time, belong to the domain of economic statics. For instance, the theory of pure exchange between buyers and sellers in a market depends upon the curves of supply and demand and consequently is a theory of economic statics. Similarly, the distribution of incomes in an economy, the number of people in each income class, displays a reasonably permanent pattern and can be described by a curve which depends little upon the variable of time.

For the purposes of this study, the definition of income as stated by Harold T. Davis in the following paragraph will be adopted:

> The nature of income is not easily attained as one may see from the constant controversy that is waged over what is to be included in income-tax returns. It will be sufficient for

> our purposes to define the income of an individual as that quantity of goods and services, measured in terms of a money unit, which he has received during some period of time as a result of the expenditure of disutility or the employment of capital during that time. It is customary to denote the first category of income as wages and salaries and the second category as the return from investment. The total income of all its citizens is known as the total income of the state (1).

This study concerns itself only with the first category of income described as wages and salaries and with reference to employee incomes in the manufacturing industries. Throughout the thesis the term income shall mean the wages and salaries of employees of the companies participating in the study from selected manufacturing industries.

The question of wages and salaries is the concern of every industrial enterprise. Employee compensation represents a substantial part of the total operating costs of any business. How well a company controls these wage costs, and how much employee productivity is obtained in return for the income dollar, can have an important effect on the success of the enterprise.

The company wage and salary program is important alike to employees, managers, and owners. The pay policies and practices of the company affect the ability of employees at every level to attain their reasonable material aspirations. To the owners of the company, wage and salary administration can be a significant means of keeping the company strong; and at the same time it influences the manager's success in his job.

Employees are sensitive about how their pay compares with that received by employees in other companies who do comparable work. Their knowledge of pay at other companies is generally quite imperfect. However, if they get the impression that their pay is lower than that in effect in other companies, serious employee unrest may result. This problem not only is an important factor in employee attitudes, but also has an important bearing on the company's ability to obtain and retain the number of types of persons required to operate efficiently. Moreover, comparative pay structures have an important effect on the company's costs and, therefore, on its ability to compete effectively in the product market.

Firms of comparable size in the same industry require personnel with similar skills and experience. Even though individual jobs and classifications will differ, the companies tend to organize their total work effort in such a way that there will be a comparable hierarchy of positions within each firm. The activity of large groups of people in relation to their wants and satisfactions is apparently not indeterminate and random. It seems reasonable to assume, therefore, that a structural relationship underlies the vagaries of individual action for the activity of the group.

The theory of economics, as it applies to a sufficiently large collection of individuals, is measurable using mathematical and statistical techniques. There exist determinate values which may be ascertained within definite limits of error. Evidence from varied sources (1), (13), and (14), leads to the conclusion that the distribution of incomes in stable societies conforms closely to the Pareto distribution of income as expressed by the formula $y = a\,x^{-v}$; y represents the number of persons receiving an income of x or more, and a and v represent the two parameters of the economic system.

So far as industrial engineering is concerned, the Pareto distribution can be usefully applied as a quantitative method of comparison in analyzing the structure of employee incomes in various industries and thus aid management in its decisions. The testing of Pareto's distribution with aggregate income has been common; but the distribution has not been extended to the study of wages and salaries in industry.

Needless to say, income is the keystone of economic planning not only at the broad aggregative level but also at the narrow individualistic level. Due attention should, therefore, be given to the wage and salary structure of the single individual enterprise. This study is devoted to this end.

CHAPTER II

SURVEY OF LITERATURE

In this chapter a background of the income distribution concept will be discussed to acquaint the reader with the subject matter. A survey of literature will be presented in order to familiarize him with the various income distribution models of significance which have been proposed up to date. Since the author has selected the Pareto income distribution model in this study, a comprehensive description of this model will be included.

Statistical data with which one works in practical problems have been classified under two types: (a) data distributed with respect to some unit of time, and (b) data distributed with respect to some physical characteristic. Most statistics used in economics are of the first type which is referred to as a <u>time series</u>. The second class is known as the <u>frequency distribution</u>, where data are distributed according to class units which are suggested by the character of the subject under observation. The individuals associated with a given class do not possess the same characteristics in the same degree. For example, classification of wage and salary earners by income is a case of frequency distribution which has the characteristics of a continuous function.

The use of mathematical methods in economics is today supplemented by a tool which was not available to such prominent economists as Alfred Marshall (1842-1924), Antoine A. Cournot (1807-1877), Lord Kelvin

(1824-1907), and Leon Walras (1834-1910). This tool is the collection of statistical data which have been assembled by many agencies since the beginning of the present century, and which affords an opportunity for the numerical verification of economic postulates. In the meantime, there has been developed a very extensive and adequate theory of statistics so that the collected data could be subjected to analysis.

A statistical description of an income distribution is defined by Aitchison and Brown (2) as a rule which gives for each value of x the proportion F(x) of persons in a given population who have an income not greater than x. Such a description, these authors further state, if it is to be a useful analytical tool, requires that F(x) be given a mathematical expression involving known, or usually unknown parameters. They both resolve on four criteria to be used in assessing the success of a particular description:

(a) To what degree does the mathematical expression approximate the observed distribution of incomes when specific values are assigned to the parameters?

(b) In arriving at the model certain assumptions are made which are consistent with the available knowledge in the field. To what extent may the statistical description be shown to rest on these assumptions?

(c) Does the statistical model provide facilities in the analysis of the data?

(d) From the economic standpoint, what meaning or significance can be attached to the parameters of the selected description?

It has been established that income distributions almost invariably possess a single mode and are positively skewed. Among the statistical descriptions which satisfy these rather general conditions and which in the past have been proposed as applicable to the distribution of incomes, works of the following authors may be mentioned: Pareto, Kapteyn and Gibrat, Frechet, Amoroso, and Champernowne.

In the year 1903 J. C. Kapteyn, in his book Skew Frequency Curves in Biology and Statistics, mentions the law of proportionate effect with reference to the distribution of income. This postulate is similar to Gibrat's (3) law of proportionate effect; the process of reasoning is from theory to observation, from the law of proportionate effect to the derivation of the lognormal description of income distribution, and then to the testing of the theory against observed data. The Kapteyn-Gibrat description is called the log-normal distribution. An application of the lognormal distribution was undertaken by H. S. Steyn (4) in studying the distribution of incomes for the Union of South Africa. Kolmogoroff (5), Halmos (6), and Epstein (7) have recently developed a new process analogous to the lognormal description for studying breakage processes and particle size distributions.

Professor Champernowne (8) describes a model which under realistic assumptions generates approximately a distribution of incomes obeying Pareto law. This model provides a basis for the comparison of processes in generating the Pareto and lognormal descriptions of income distribution. His model assumes, in a grouped form, equal probabilities for proportionate changes in income.

The best known attempts to derive Pareto's empirical income-description from simple assumptions are those of Rhodes (9) and Castellani (10). Rhodes makes assumptions concerning the distribution in the population of certain talents which are held to determine the income-earning capacity of an individual in conjunction with a socially accepted system of rewards. It also involves "a certain mathematical assumption" for which he asserts there is "no justification apart from expediency."

Castellani's model is based on a physical analogy taken from the field of statistical mechanics. He first divides the income range into intervals, and then determines the number of incomes falling into each interval by a multinomial process. The assumption he makes is that there exists independent probabilities associated with each interval and that something akin to physical forces of attraction constrains the operation of the process.

The first extensive study from the statistical point of view on the problem of how income is distributed among the citizens of a state was carried out by Vilfredo Pareto (1848-1923). Pareto was a disciple of Leon Walras and his successor in the chair of Political Economy at the University of Lausanne. In 1897 publication of his second book entitled Cours d'economie politique marks the beginning of the study of the income distribution problem. The first chapter of Pareto's book is entirely devoted to this problem (11).

Upon collection of income data from numerous sources and employing statistical tools, Pareto arrived at a formulation of his law of income distribution. This law has been described by Professor Harold T. Davis somewhat precisely in the following statement (1):

> In all places and at all times the distribution of income in a stable economy, when the origin of measurement is at a sufficiently high income level, will be given approximately by the empirical formula $y = a\,x^{-v}$, where y is the number of people having the income x or greater, and v is approximately 1.5.

The significance of the Pareto law is very well reflected in the comments of Professor Snyder (12):

> But there is a larger interest in this very interesting field and that is whether the Pareto curve is not a general expression for practically every kind of highly specialized ability, from that of the billiard player or baseball pitcher to the spectacular performance of the chess player, . . . Clearly there are not in this world many Newtons or Shakespeares or Faradays, any more than there is a large population of Dizzy Deans or Willie Hoppes. . . . Such at least was the conception to which the writer had given a good deal of attention. . . . Enough, however, it has seemed to me, to indicate that the thesis is valid, and further that the Pareto Curve is destined to take its place as one of the great generalizations of human knowledge.

Keeping in mind the criteria used by Aitchison and Brown in formulating the statistical description of income, the mathematics of the Pareto distribution will be presented next based on Professor Davis' presentation. Let us assume a population of N individuals who are to be distributed with respect to their possession of a quantity of a variable x; let the distribution function be designated by Z(x) and furthermore, let the lowest measure of the range of x be A and the highest B. In other words, $A \leq x \leq B$ as indicated in Figure 1 below.

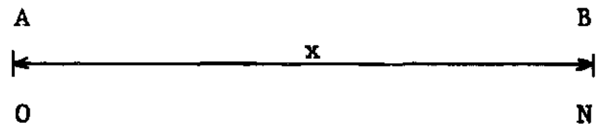

Figure 1. Range of the Distribution Function

The total number of individuals who possess the measure between x and x + dx was defined by Z(x). For the range from A to X, the number of individuals falling in this category can be described by the integral

$$Y(X) = \int_{A}^{X} Z(x)\ dx. \tag{1}$$

Then the number of individuals who have the measure from A to B is given by the relation

$$Y(B) = \int_{A}^{B} Z(x)\ dx = N. \tag{2}$$

However, in defining the Pareto distribution it was stated that the relation $y = ax^{-v}$ holds when the origin of measurement is at a sufficiently high income level (e.g., the relation describes best the range from X to B). Thus, N - Y(X) must be the frequency function of the population in question, since it is this function which gives the number of individuals who have the measure X or greater.

If the cumulative frequency function is designated by y(X), the following relation must hold

$$y(X) = N - Y(X) = \int_X^B Z(x)\ dx. \tag{3}$$

This is the function which was assumed to have the form $y = a\ x^{-v}$ by Pareto. If we accept Pareto's proposition, the following simple mathematical relations can be obtained from the above function.

The cumulative frequency function is assumed to have the form

$$y = a\ x^{-v}. \tag{4}$$

Differentiating equation (4), y with respect to x, we get

$$\frac{dy}{dx} = (- v\ a\ x^{-v-1} + 0) = Z(x).$$

Simplifying, $$Z(x) = - \frac{dy}{dx} = (v\ a)\ x^{-v-1}. \tag{5}$$

Taking logarithms of equations (4) and (5) the linear relationships below are obtained:

$$\log y = \log a - v \log x, \tag{6}$$

$$\log z = \log (av) - (v+1) \log x. \tag{7}$$

When the logarithms of the two functions above versus the logarithms of x are graphed on full-logarithmic paper, the graphs will be two straight lines with slopes of v and (v+1) respectively. The Pareto curve in its more general form is expressed in equation (8),

$$\log N = \log A - \alpha \log (a+X) - \beta X. \tag{8}$$

Equation (6) represents only a simplification of this more general equation. In (8) β was so small that Pareto neglected it thus reducing equation (8) to the form

$$\log N = \log A - \alpha \log (X+a), \tag{9}$$

where N is the number of individuals with income equal to, or greater than X; A, α, a are constants. The expression in (6) has been referred to as the first law, while equation (9) is known as his second law of income distribution. In this study, we will be concerned with the application of Pareto's first law which seems to be the most widely known formula.

Many attempts have been made to obtain data on the incomes of individuals from the lowest income levels to the highest. The most comprehensive study is that made by the National Bureau of Economic Research, who published their estimates of the distribution of income among personal-income recipients in the United States for the year 1918. Davis shows the fantastic spread of the income from the average by computing the second, third, and fourth moments about the mean, the unit being $1000 for the above U. S. data. The calculated values were

$\mu_2 = 32.1367, \mu_3 = 40165.4694, \mu_4 = 77,281,288.7$. From these moments Professor Davis computes the standard error, the skewness, and the kurtosis of the distribution to be respectively $\sigma = 5.6689$, $S = \mu_3/(2\sigma^3) = 11.0253$, $E = \beta_2 - 3 = 74,826$, where β_2 is taken to be (μ_4/σ^4).

The foregoing investigations of personal-income recipients in the U. S. for the year 1918 led Davis to the conclusion below:

> These values show how hopeless is the task of attempting to graduate the data by any of the curves of Pearson type. This comment is quite significant, since the problem invoked by the distribution of incomes is thus shown to be essentially different from that of the usual frequency distributions which arise in biology for which the Pearson types were primarily designed. The extraordinary difference between biometric frequencies and income frequencies is found in the general observation that in the former, even in cases of extreme skewness, it is unusual to find any member of the distribution more than 4σ from the mean. In the case of income data, extreme individuals are found more than 700σ from the mean (1).

A verification of the general thesis of the Pareto distribution of special abilities is given in the data published in the 28th annual report of the General Motors Corporation for the year ending December 31, 1936. The report shows the salary schedule for the executive administrative staff of the corporation. Having arranged the data from this report in a cumulative frequency table, Davis fits a parabolic curve to the data using the least squares method. The equation

$$\log y = 10.83006 - 2.067 \log x \qquad (10)$$

was obtained and Davis' interpretation of his results is that the distribution of salaries in a successful corporation like General Motors tends toward the Pareto pattern.

In 1951, Hayakawa (13) attempted to estimate income distribution as a whole for Japan, using data obtained from 108 communes (villages, towns, and cities) and applying the Pareto curve. Statistics of incomes in Japan indicated that the Pareto curve holds good for the upper range of the income distribution and that such an extension to the lower income ranges is not warranted.

Professor Shirras (14) analyzed the distribution of income in India for the period 1913-1914 to 1929-1930 using the Pareto distribution. In 1935, his conclusion was that the logarithmic curve of the distribution of incomes did not fall into a straight line and that the law did not hold at all. Twenty-one years after Shirras, however, Bose and Roy (15), in view of the changed economic conditions, better coverage, and the availability of improved statistical data, attempted to test the validity of the Pareto hypothesis in India. Their results and conclusions were in support of the validity of the Pareto hypothesis.

In England, Professor Allen (16) made an estimate of the distribution of higher salaries of men in the United Kingdom for the period 1954-1955. When his cumulative frequencies are plotted on a full-logarithmic paper against income, they fall close to a straight line with slope of about -2.5. One of the United Nations publications (17) gives an estimate of the distribution of wage and salary income in Poland for September, 1955. The value of the slope was found to be 4.5.

Among the latest works in the field of income distribution are those of Lee Soltow (18) and Foster, Wajda, Lawson (19). Soltow's study is designed to investigate changes in the dispersion of wages and salaries in one manufacturing plant over a long period of time. Foster, Wajda, and Lawson propose a method which they describe as the "Global Plan," and by which companies can attempt to establish how their salary schedule compares with the schedule of other companies within an industry. Even though the basic objective seems to coincide with that of the author, the "Global Plan" approach has left out the wage structure problem entirely.

The National Industrial Conference Board (20), a non-profit institution for scientific research in the fields of business economics and management, has published a series of reports on the subject of compensation of top executives in various industries. In these reports, the relationship of executive pay to a company's sales volume is emphasized which is then followed by a discussion of the ratios of base salary represented by incentive and bonus payments. The Bureau of Labor Statistics of the United States Department of Labor can be mentioned as another institution engaged in the study of wage and salary distributions.

The above literature survey of the distribution of incomes leads the author to the conclusions below:

In studying the distribution of incomes the well known procedure has been to compare actual observations with a mathematical model. Although such a model is never perfect, many valuable conclusions can be drawn provided the model chosen gives an adequate fit to a set of observations. If the model selected fails to fit the observations sufficiently well, there is the possibility of modifying or even changing the model to obtain a better fit.

The forces determining the distribution of incomes are so varied and complex, and interact and fluctuate continuously, that any theoretical model must be either simplified or hopelessly complicated. The general procedure has been to choose a simple model which can be easily interpreted rather than selecting a more complex model which, while it gives a better fit, is more difficult to apply and interpret. In this study, therefore, the author has selected the model proposed by Pareto in attempting to analyze the structure of wages and salaries in the various industries.

CHAPTER III

PROCEDURES

When studying the association between two or more variables, it is necessary to base the study on various assumptions. Although these assumptions may be shown to be reasonable and not inconsistent with the observations, it is, in general, impossible to prove any assumptions that are made. In order to limit the practical considerations of the wage and salary distribution problem, the assumptions on the basis of which the observed data will be analyzed can be listed as follows:

(1) The theory of economics, as it applies to sufficiently large collection of individuals, is measurable using mathematical and statistical techniques. There exist determinate values which may be ascertained within definite limits of error.

(2) The activity of large groups of people in relation to their wants and satisfactions is not random and indeterminate. There is a trend and structural relationship for the activity of the group.

(3) Firms of comparable size in the same industry require personnel with similar skills and experience. Even though individual jobs and classifications will differ, the companies tend to organize their total work effort in such a way that there will be a comparable hierarchy of positions within each firm.

(4) The basic frequency function as applicable to (2) and (3) above can be described by the Pareto formula $y = a\,x^{-v}$, where y

represents the number of persons receiving an income of x or more, a and v represent the two parameters to be determined for each firm participating in the study.

(5) Income data, obtained by confidential questionnaires from individual establishments in selected manufacturing industries, are such as to reflect the pattern of income distribution being representative of the situation they refer to.

(6) In fitting a line to collected data, the deviations from the fitted line $\log y = \log a - v \log x$ are normally distributed with the same variability. The unaccountable variability can be attributed to a number of independent causes, each of which contributes a part of the resulting deviation.

This survey of employee earnings in selected manufacturing industries represents establishments having 100 or more paid employees. The study is limited to participating firms in 14 manufacturing industry groups selected from Plant and Product Directory (21) which gives a list showing the employment, industry classification (SIC), and location of the individual manufacturing establishments.

The term establishment, for the purposes of this study, shall mean a single physical location where industrial operations are performed. It is not necessarily identical with the company, which may consist of one or more establishments.

Data used in the study were gathered by mail questionnaires. A total of 211 confidential questionnaires were mailed to the Wage and Salary Administrator of the establishments included in the study. Participating firms were requested to furnish monthly wage and salary data on seven categories of employees. The request letter directed to the Wage and Salary Administrator of each company, the instructions for filling out the questionnaires, and the questionnaire are shown on pages 20, 21, and 22, respectively.

Replies were received from 19 establishments of which four were excluded from the study because the data were incomplete. The calculations are based on 15 establishments. Table 1 on page 24 gives a tabulation of the number of establishments contacted in the 14 industry groups, the SIC code, the product classification, and the number of participants from each group. Should the reader be interested, excerpts from reply letters of the participating as well as the non-participating establishments are illustrated on page 89 in the Appendix.

The employment positions (see "Instructions" sheet) of a firm were divided into seven payroll categories and income data were requested on each of these seven. Originally, the intent was also to include a comparative study among the seven payroll categories within each company. However, the numbers of persons corresponding to several of the payroll categories were so small that it was not feasible to analyze the data based on payroll classification. It was decided, therefore, to combine income data of the seven payroll categories in obtaining the income data for a particular company.

A tabulation of the wage and salary data as furnished by individual companies is presented in Tables 2 through 4 of this chapter. Capital letters A through O were assigned to identify the respective companies. Then, tables of cumulative frequencies were formed from the collected data of the participating establishments. These cumulative

GEORGIA INSTITUTE OF TECHNOLOGY
Atlanta 13, Georgia

School of
Industrial Engineering

November 5, 1962

The X Y Z Company
Wage and Salary Administrator
762 West Kessler Blvd.
Indianapolis 5, Indiana

Gentlemen:

I am writing with reference to a research project leading to the MS degree and carried out under the direction of Dr. Joseph Krol, Professor of the School of Industrial Engineering. My field of interest is to analyze the structure of wages and salaries in some selected industries. In one phase of the project, I would like to establish whether the Pareto distribution could be usefully applied as a quantitative method of comparison between various industries and individual companies.

An extensive literature survey investigating the basic data required in the project has been completed. However, some additional data, not available in the Atlanta area, is needed. It would be greatly appreciated if you would furnish the requested data as indicated on the enclosed questionnaire forms. A list of specific requirements is also enclosed. Your cooperation in filling out the questionnaires will definitely contribute to the success of this project.

You can rest assured that all information furnished by your organization will be held in confidence. The code names A, B, C, etc., will be used with reference to the participating companies for the purposes of this study.

Thank you for your consideration in this matter.

Very truly yours,

Apel Dido, Graduate Student
School of Industrial Engineering

Encls: 8

I N S T R U C T I O N S

PLEASE NOTE: All information furnished by your organization will be held STRICTLY CONFIDENTIAL.

I. The employment positions of your firm has been divided into seven payroll categories for the purposes of this study. You are requested to furnish monthly salary and wage data on the following categories and ONLY IN COLUMN (A) of the questionaire.

1. Scientific - Professional Nonsupervisors
2. Scientific - Professional Supervisors
3. Administrative Nonsupervisors
4. Administrative Supervisors
5. Production - Workers Nonsupervisors
6. Production - Workers Supervisors
7. All Others

II. For survey purposes, supervisory shall mean those directing the activities of five or more permanently assigned employees. Administrative will cover all those who are exempt under the Fair Labor Standards Act and who are performing in positions considered as other than scientific-professional. Employees such as engineers, physicists, and mathematicians would most certainly be scientific professional, while those engaged in personnel or industrial relations, accounting, purchasing, maintenance, and similar occupations would be categorized as administrative. The final decision as to which category to use for an individual employee or a particular group of employees will be left to your organization.

III. Please indicate the total number of people employed by your company on any one of the seven enclosed questionnaire forms in the space provided at the bottom of the page. Your cooperation in completing the questionnaires as soon as possible will be greatly appreciated. We would like to have the requested data by ________________________ if possible at all. Thank you very much.

RETURN ADDRESS: Dr. Joseph Krol, Professor
School of Industrial Engineering
Georgia Institute of Technology
Atlanta 13, Georgia

SALARY & WAGE INFORMATION QUESTIONNAIRE

COMPANY: ______________________ DATE: ______________________

PAYROLL CATEGORY: ______________ DATE AS OF: ______________
(date)

DISTRIBUTION OF BASE MONTHLY SALARIES BY PAYROLL CATEGORY

MONTHLY SALARY INTERVAL	(A) NUMBER OF PERSONS	PERCENT OF COLUMN (A) TOTAL	MONTHLY SALARY INTERVAL	(A) NUMBER OF PERSONS	PERCENT OF COLUMN (A) TOTAL
Under $200			2500 - 2999		
$200 - 299			3000 - 3499		
300 - 399			3500 - 3999		
400 - 499			4000 - 4499		
500 - 599			4500 - 4999		
600 - 699			5000 - 5499		
1600 - 1699			10000 - 12499		
1700 - 1799			12500 - 14999		
1800 - 1899			15000 - 17499		
1900 - 1999			17500 - 19999		
2000 - 2499			20000 & Over		
			TOTAL		100 %

TOTAL SUM OF ALL THE EMPLOYEES
IN YOUR COMPANY: ______________________

frequency tables, Table 16 through Table 34, can be seen in the Appendix. It should be noted from the cumulative frequency tables that the class mark (X) was arbitrarily chosen as the center of the class interval.

Employing the method of least squares, the following normal equations were formed for computing the parameters log a and v in equation (6):

$$N \log a - v \sum \log x = \sum \log y \tag{11}$$

$$\sum (\log x) \log a - v \sum (\log x)^2 = \sum (\log x) (\log y). \tag{12}$$

The values of N, $\sum \log x$, $\sum \log y$, $\sum (\log x)^2$, and $\sum (\log x)(\log y)$ were taken from the frequency table of the company being analyzed. The constant N represents the total number of class marks in the company's income structure. Once computed, values of log a and v were substituted in equation (6) giving us the desired curve for the particular company, in logarithmic form.

The 15 individual companies were subjected to statistical procedures as outlined above. In arriving at the desired curve for a given industry, data of all the establishments within the same industry were combined together. To indicate the closeness with which the observed cumulative frequency values is represented by the computed values, the computed values are presented in the cumulative frequency tables. Both the observed and the computed values are graphically represented on full logarithmic paper in Figures 2 through 20 of Chapter IV.

Table 1. Number of Participating Companies Out of Those Contacted by Confidential Questionnaires in the Various Industries

SIC CODE	INDUSTRY GROUP BY PRODUCT CLASSIFICATION	NUMBER OF ESTABLISHMENTS CONTACTED	NUMBER OF PARTICIPANTS IN EACH INDUSTRY
2111	Cigarettes	12	None
2211	Broad-woven Fabric Mills, Cotton	36	2 (1 Unusable)
2311	Men's, Youth's & Boy's Suits, Coats & Overcoats	17	None
2911	Petroleum Refining	12	2
3011	Tires & Inner Tubes	5	None
3334	Primary Production of Aluminum	28	8 (2 Unusable)
3519	Internal Combustion Engines	5	None
3571	Computing & Accounting Machines	10	3
3651	Radio & Television Receiving Sets	5	None
3661	Telephone & Telegraph Apparatus	5	None
3721	Aircraft	53	4 (1 Unusable)
3731	Ship Building & Repairing	5	None
3732	Boat Building & Repairing	5	None
3861	Photographic Equipment & Supplies	8	None
TOTAL	14	211	19

NOTE: Income data of four companies were excluded from this study because the data were incomplete.

Table 2. Questionnaire Data as Furnished by the Participating Companies A Through H

MONTHLY SALARY INTERVAL	NUMBER OF PERSONS (Y) IN COMPANY							
	A	B	C	D	E	F	G	H
$200 - 299	0	0	0	0	3	0	7	0
300 - 399	13	7	0	2	329	2	51	12
400 - 499	747	983	38	45	737	19	588	574
500 - 599	1,622	751	199	568	174	29	2,537	338
600 - 699	234	272	33	305	53	47	269	2,653
700 - 799	141	147	10	99	45	25	333	144
800 - 899	110	84	11	20	25	10	206	218
900 - 999	51	36	5	6	12	4	122	136
1000 - 1099	27	17	0	3	5	2	85	76
1100 - 1199	1	5	0	3	2	1	75	64
1200 - 1299	0	3	0	1	1	0	49	40
1300 - 1399	0	0	0	0	2	0	46	13
1400 - 1499	0	0	0	0	1	0	22	17
1500 - 1599	0	0	0	1	6	0	13	11
1600 - 1699	0	0	0	0	0	0	2	4
1700 - 1799	0	0	0	0	1	0	0	1
2000 - 2499	0	0	0	1	1	0	0	0
20,000 & Over	0	0	0	0	0	0	35	0
Totals	2,946	2,305	296	1,054	1,397	139	4,440	4,301

Table 3. Questionnaire Data as Furnished by the Participating Companies I Through O

MONTHLY SALARY INTERVAL	NUMBER OF PERSONS (Y) IN COMPANY						
	I	J	K	L	M	N	O
$200 - 299	0	0	48	53	0	734	5,948
300 - 399	1,020	0	58	182	363	316	2,936
400 - 499	2,464	2	1,203	2,832	1,684	406	349
500 - 599	1,067	136	235	1,168	3,341	258	229
600 - 699	213	162	52	487	1,804	122	72
700 - 799	200	177	24	325	925	57	39
800 - 899	137	143	15	447	761	29	28
900 - 999	149	131	4	104	545	7	16
1,000 - 1,099	41	127	2	103	361	2	8
1,100 - 1,199	17	88	0	46	203	0	6
1,200 - 1,299	24	65	0	55	144	3	9
1,300 - 1,399	8	41	0	18	93	2	6
1,400 - 1,499	0	30	0	6	46	0	2
1,500 - 1,599	1	19	0	11	49	0	6
1,600 - 1,699	6	5	0	11	24	0	1
1,700 - 1,799	6	1	0	0	0	0	5
1,800 - 1,899	3	0	0	4	0	2	0
1,900 - 1,999	2	0	0	3	0	0	1
2,000 - 2,499	5	1	0	1	0	1	1
2,500 - 2,999	2	0	0	0	0	0	3
3,000 - 3,499	1	0	0	0	0	1	0
3,500 - 3,599	2	0	0	0	0	0	3

Table 3. Questionnaire Data as Furnished by the Participating Companies I Through O (Continued)

MONTHLY SALARY INTERVAL	NUMBER OF PERSONS (Y) IN COMPANY						
	I	J	K	L	M	N	O
5,000 - 5,499	2	0	0	0	0	0	1
7,000 - 7,499	1	0	0	0	0	0	0
8,000 - 8,999	0	0	0	0	0	0	1
10,000 - 12,499	1	0	0	0	0	0	0
Totals	5,372	1,128	1,641	5,856	10,343	1,940	9,670

Table 4. Questionnaire Data Compiled into Four Industry Groups

MONTHLY SALARY INTERVAL	NUMBER OF PERSONS (Y) IN INDUSTRY			
	I	II	III	IV
$200 - 299	3	7	48	787
300 - 399	353	63	1,078	861
400 - 499	2,569	1,162	3,669	4,922
500 - 599	3,343	2,875	1,438	4,767
600 - 699	949	2,922	427	2,413
700 - 799	467	477	401	1,307
800 - 899	260	424	295	1,237
900 - 999	114	258	284	656
1,000 - 1,099	54	161	170	466
1,100 - 1,199	12	139	105	249
1,200 - 1,299	5	89	89	202

Table 4. Questionnaire Data Compiled into Four Industry Groups (Continued)

MONTHLY SALARY INTERVAL	NUMBER OF PERSONS (Y) IN INDUSTRY			
	I	II	III	IV
1,300 - 1,399	2	59	49	113
1,400 - 1,499	1	39	30	52
1,500 - 1,599	7	24	20	60
1,600 - 1,699	0	6	11	35
1,700 - 1,799	1	1	7	0
1,800 - 1,899	0	0	3	6
1,900 - 1,999	0	0	2	3
2,000 - 2,499	2	0	6	2
2,500 - 2,999	0	0	2	0
3,000 - 3,499	0	0	1	1
3,500 - 3,999	0	0	2	0
4,000 - 4,499	0	0	0	0
4,500 - 4,999	0	0	0	0
5,000 - 5,499	0	0	2	0
5,500 - 5,999	0	0	0	0
6,000 - 6,499	0	0	0	0
6,500 - 6,999	0	0	0	0
7,000 - 7,499	0	0	1	0
7,500 - 7,999	0	0	0	0
8,000 - 8,999	0	0	0	0
9,000 - 9,499	0	0	0	0
9,500 - 9,999	0	0	0	0

Table 4. Questionnaire Data Compiled into Four Industry Groups (Continued)

MONTHLY SALARY INTERVAL	NUMBER OF PERSONS (Y) IN INDUSTRY I	II	III	IV
10,000 - 12,499	0	0	1	0
20,000 & Over	0	35	0	0
Totals	8,137	8,706	10,139	18,139

Analysis of observed and computed values in Chapter IV revealed that, in general, actual income distributions of the participating companies did not adequately follow Pareto's law over the entire range of incomes. It was decided, therefore, to investigate the possibilities of modifying the adopted distribution function $y = a\ x^{-v}$.

Chapter V discusses three basic approaches undertaken by the author in attempts to describe the entire income range either by a new function, or by the combination of two independent functions. Having selected the method most feasible, data of three individual companies and one industry group were analyzed and the results presented in tabular form.

CHAPTER IV

RESULTS

In every act of inference we are concerned with a certain identity, sameness, and resemblance apparent between two objects. The principal value of science is the power to apply to one object the knowledge acquired from like objects. This is done by means of the logic of deduction and the ensuing inference. It is only so far as we can discover and register similarities that we can impute value to our observations. In proportion as resemblance is deeper and general, the commanding powers of inference become stronger. Thus, inference is the bridge by which we pass from case to case. It is the purpose of this chapter to trace out the various characteristics which the income data of the surveyed companies indicate and the possible generalizations of these characteristics.

When the data have been collected and suitably arranged in tabular form, the next step in statistical procedure is to represent them by some graphical method. Tables 16 through 34 in the Appendix show the closeness of agreement between the observed and calculated values. To illustrate the closeness of fit between the observed values of y and the computed values of y, these values are plotted on full logarithmic paper. Figures 2 through 20 of this chapter graphically portray the observed and computed income data for each company and industry group.

We observe that the cumulative income distribution curves of the companies and the industry groups reveal five basic features in common. The inherent characteristics of the curves can be summed up as follows:

(a) The general shape of the curves on full logarithmic paper, both for the companies and industry groups, is that of an inverted J.

(b) Their slopes are steeper as compared with the Paretian slope which has an average value of - 1.5; the computed values of the slopes for the income data varied between - 2.06165 and - 5.75101 (See Table 5).

(c) Income data in the upper salary range (R_u) and corresponding to the lower section of the inverted J fell on a straight-line with few exceptions. This is indicative of the fact that the portion of the curve representing the distribution of incomes in range R_u is in compliance with the general thesis of Pareto.

(d) Income data in the lower salary range (R_1) and corresponding to the curved portion of the inverted J failed to follow the general pattern of Pareto's formula $y = a\, x^{-v}$. This behavior would suggest that the Pareto distribution does not apply to the major portion of range R_1.

(e) When extreme values of the income data, especially in range R_1, are excluded from the calculations the general straight-line fit to the remaining portion of the cumulative frequency curve is appreciably improved.

A summary of the calculated Paretian constants in the general equation $y = a\, x^{-v}$ is presented in Table 5 for the individual companies and the industry groups. The first set of values corresponding to a

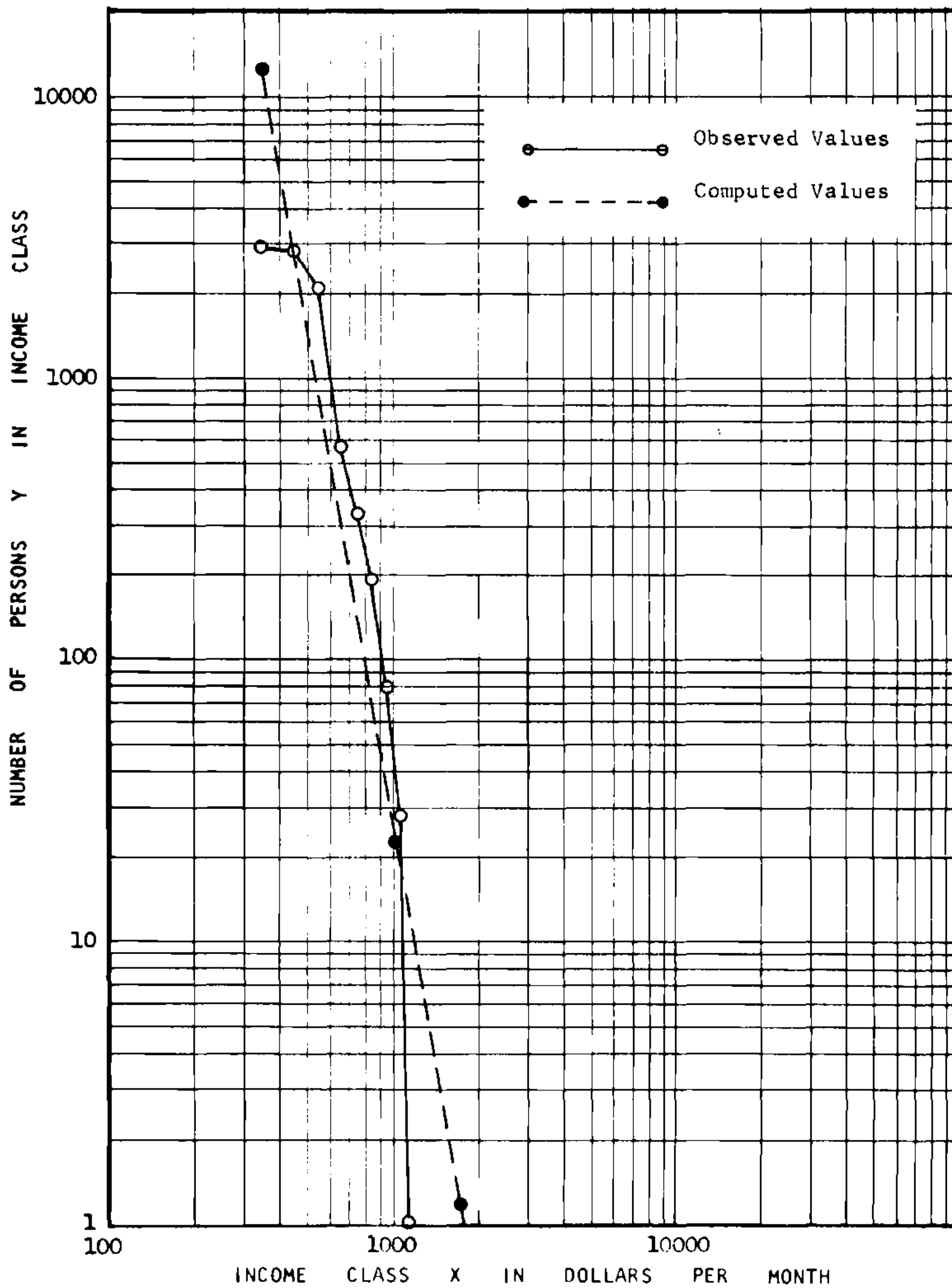

Figure 2. Wage and Salary Distribution of Company A

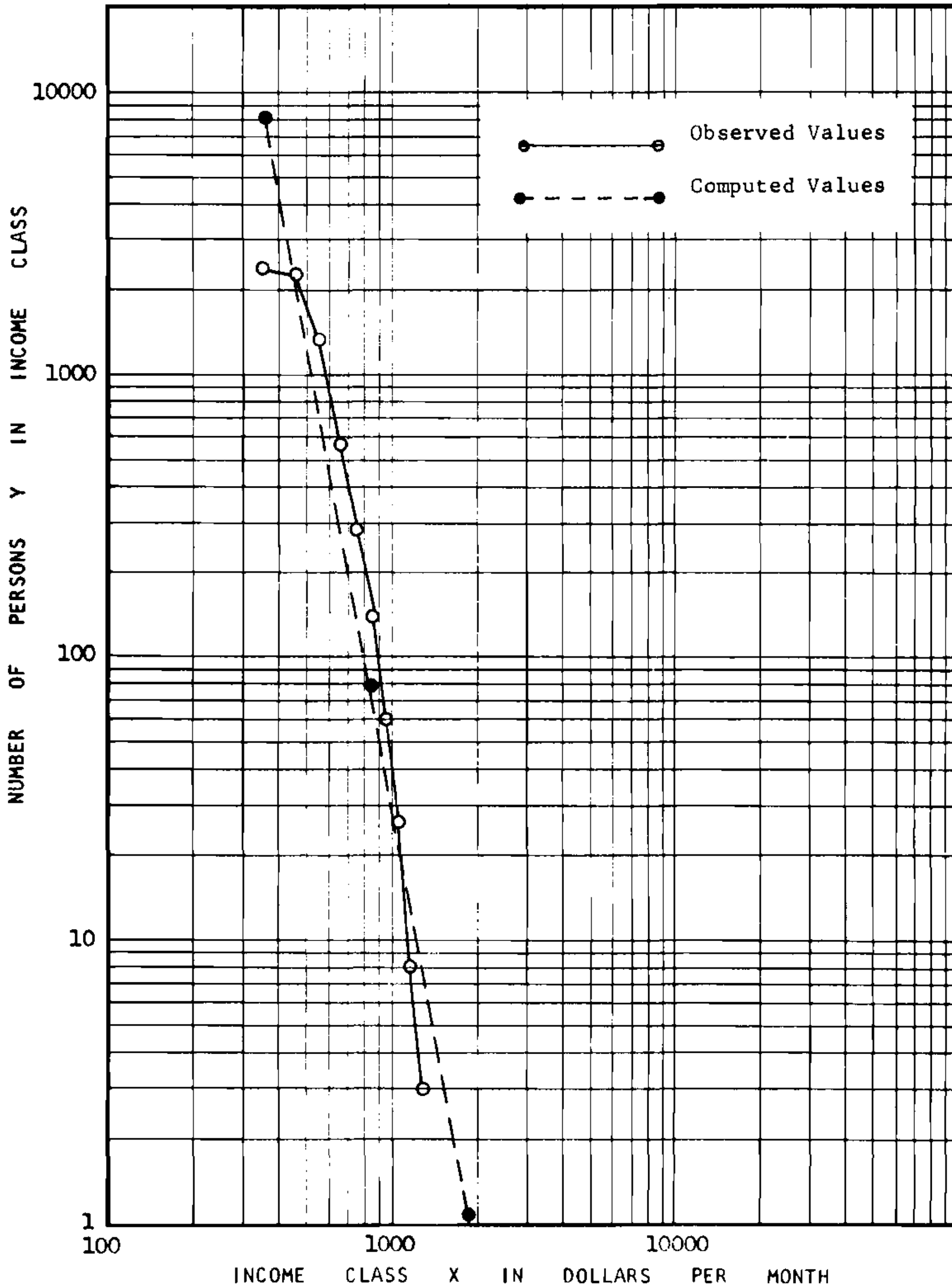

Figure 3. Wage and Salary Distribution of Company B

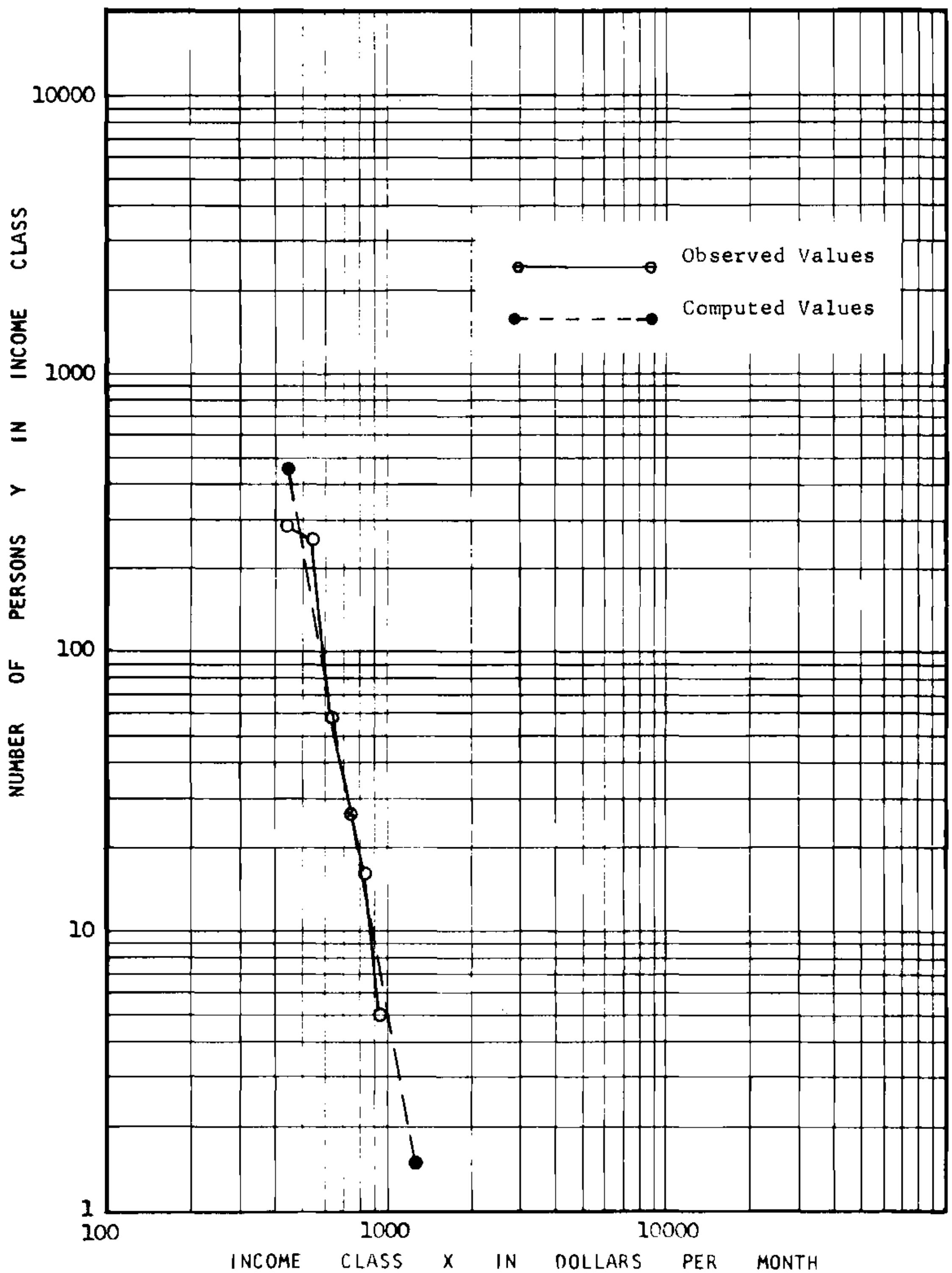

Figure 4. Wage and Salary Distribution of Company C

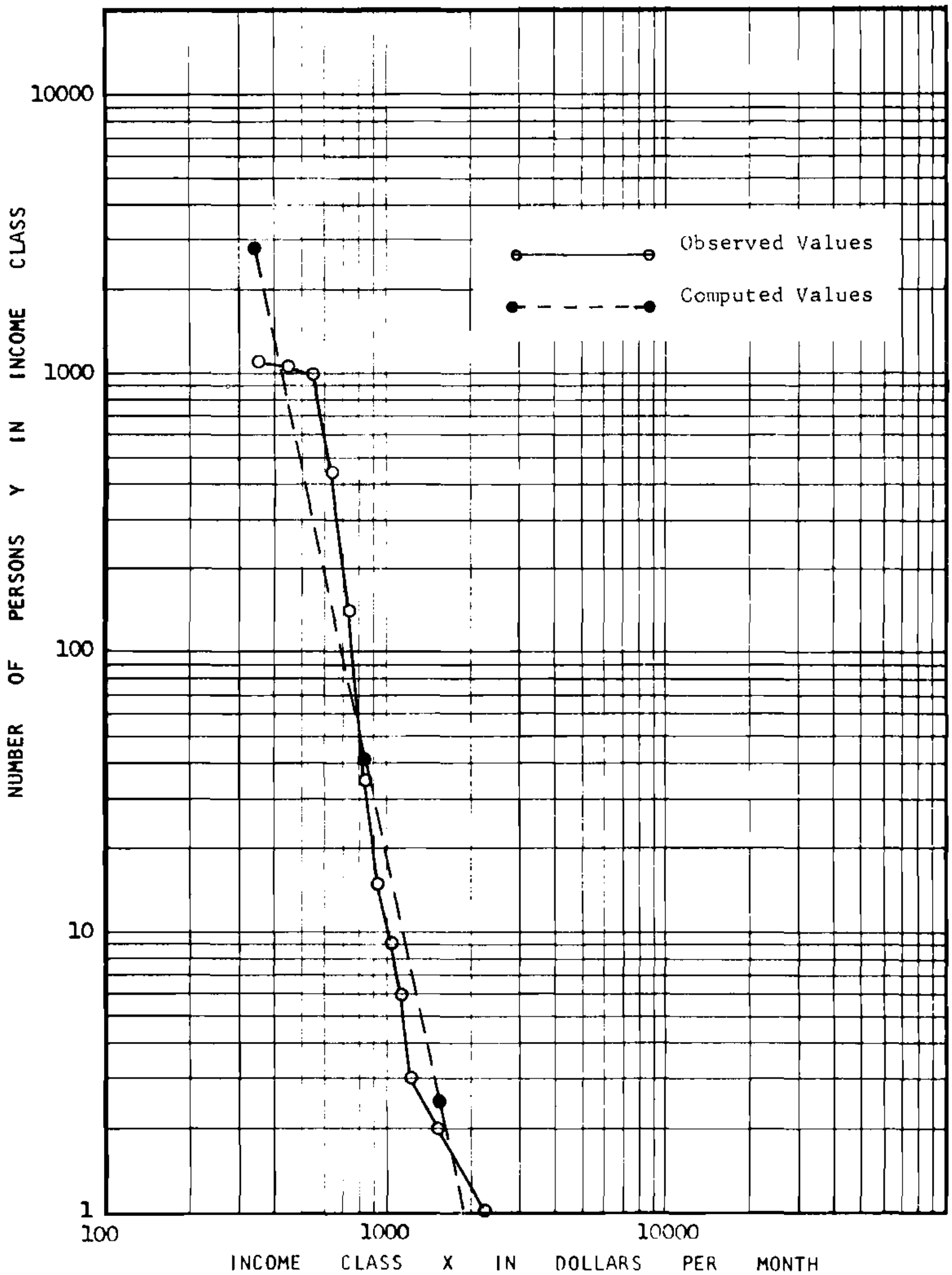

Figure 5. Wage and Salary Distribution of Company D

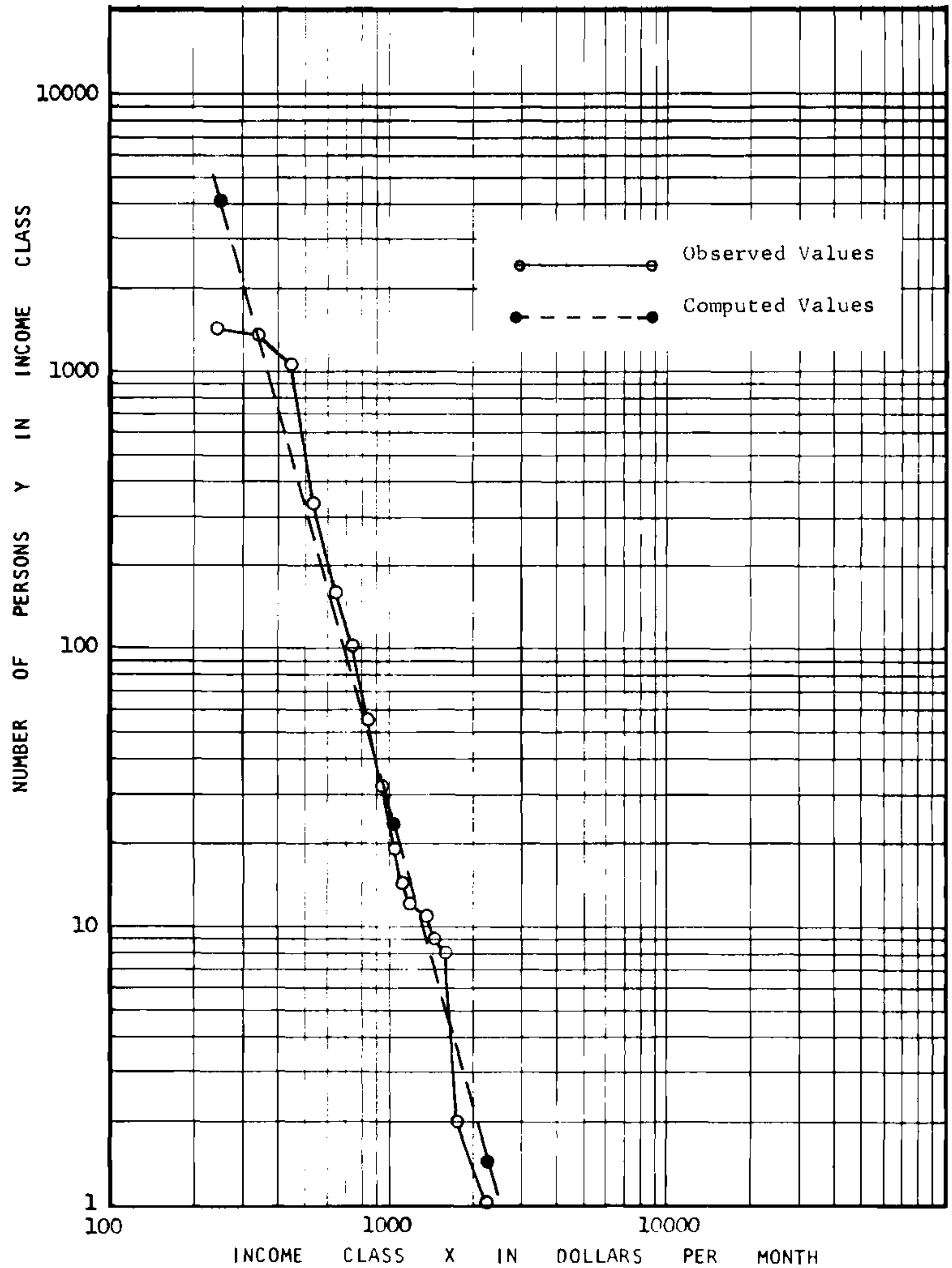

Figure 6. Wage and Salary Distribution of Company E

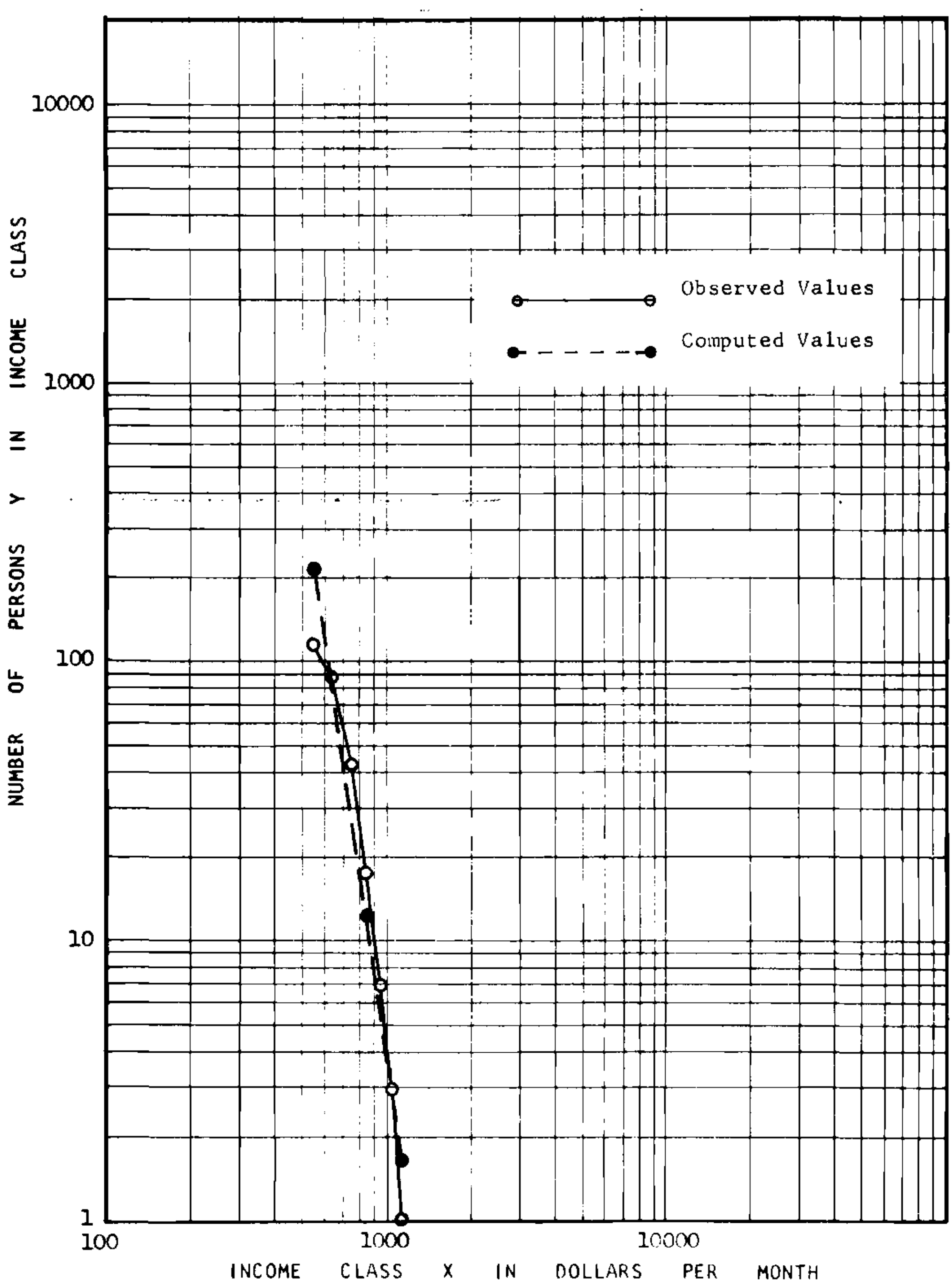

Figure 7. Wage and Salary Distribution of Company F

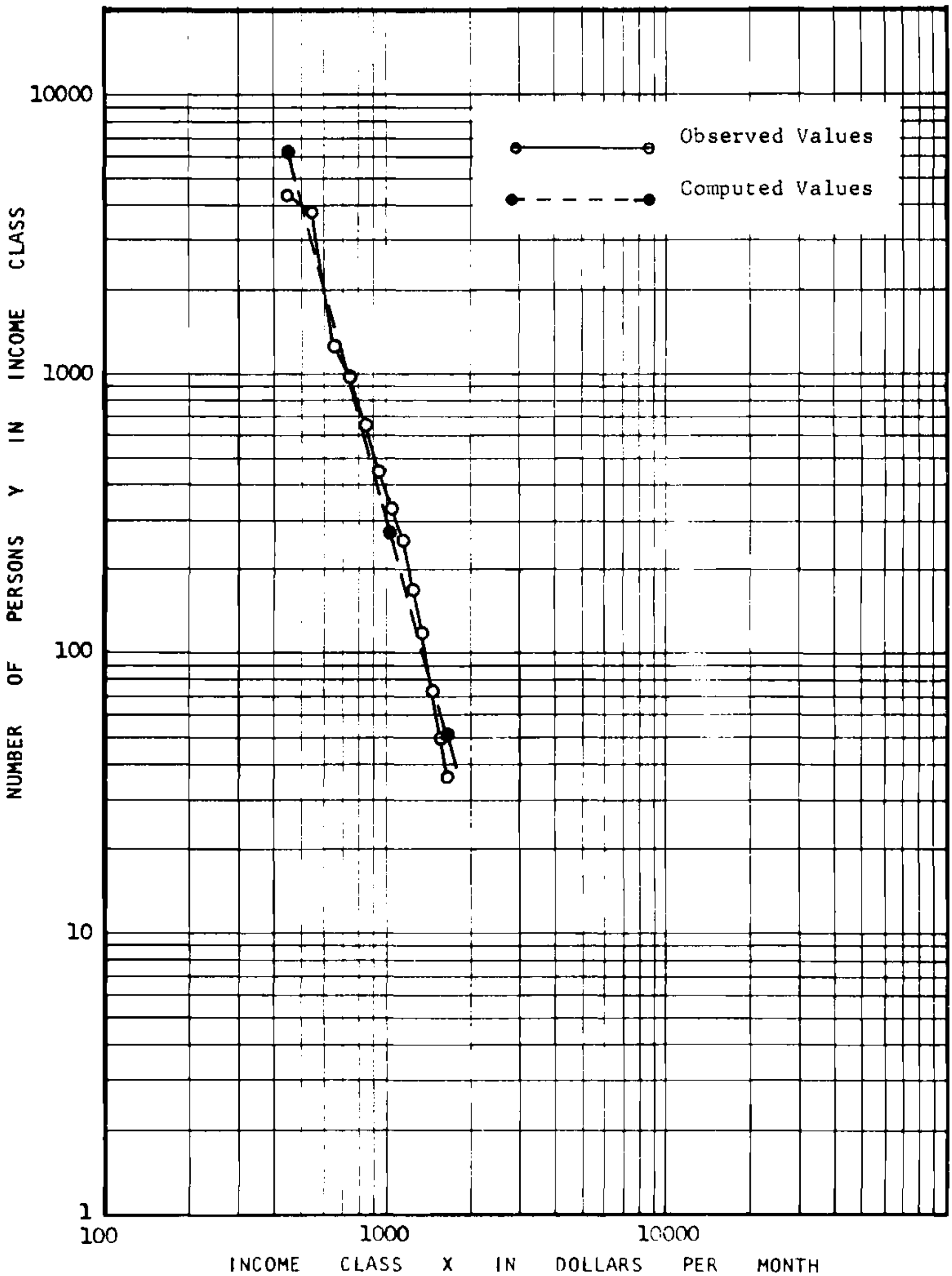

Figure 8. Wage and Salary Distribution of Company G

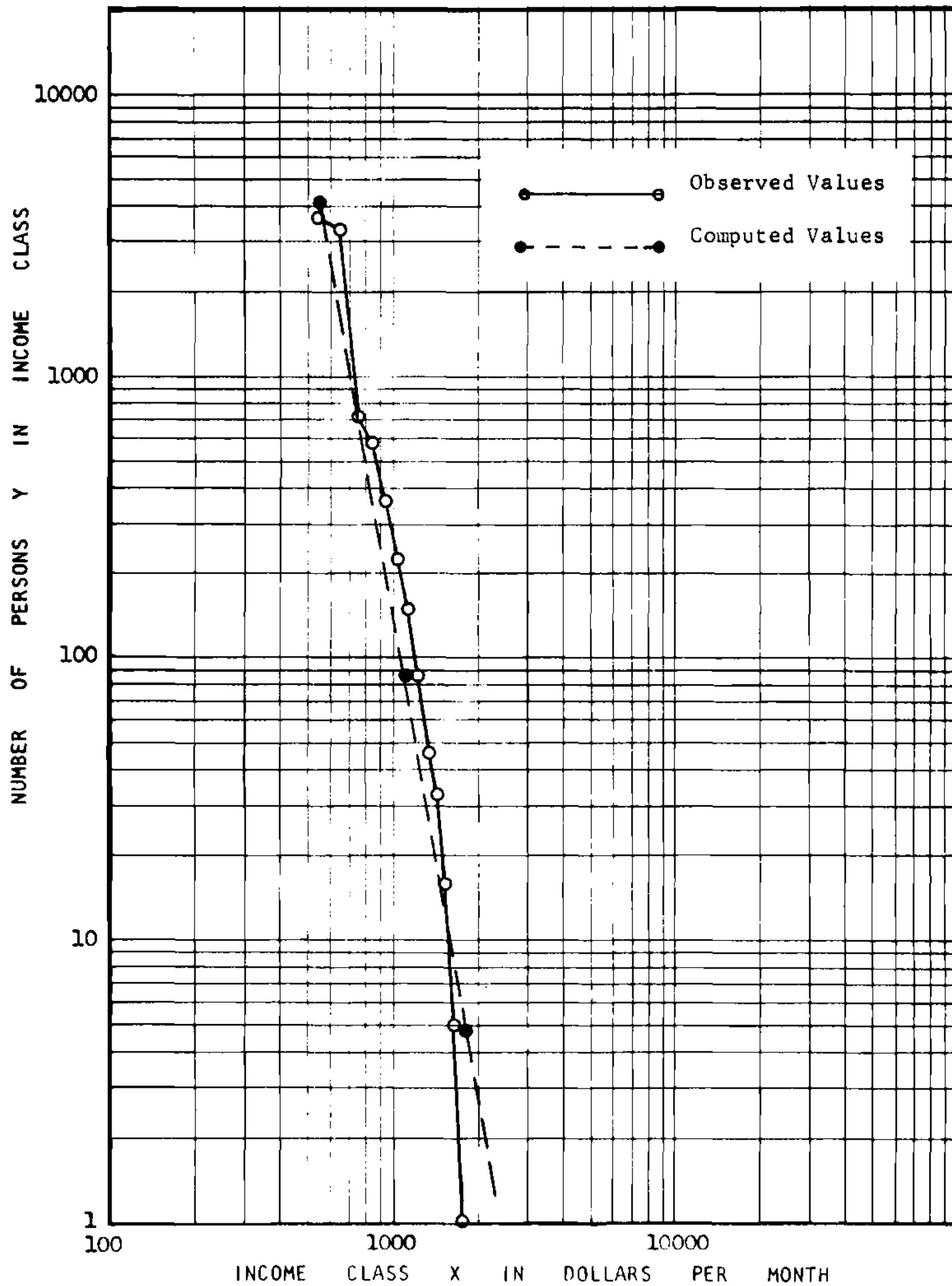

Figure 9. Wage and Salary Distribution of Company H

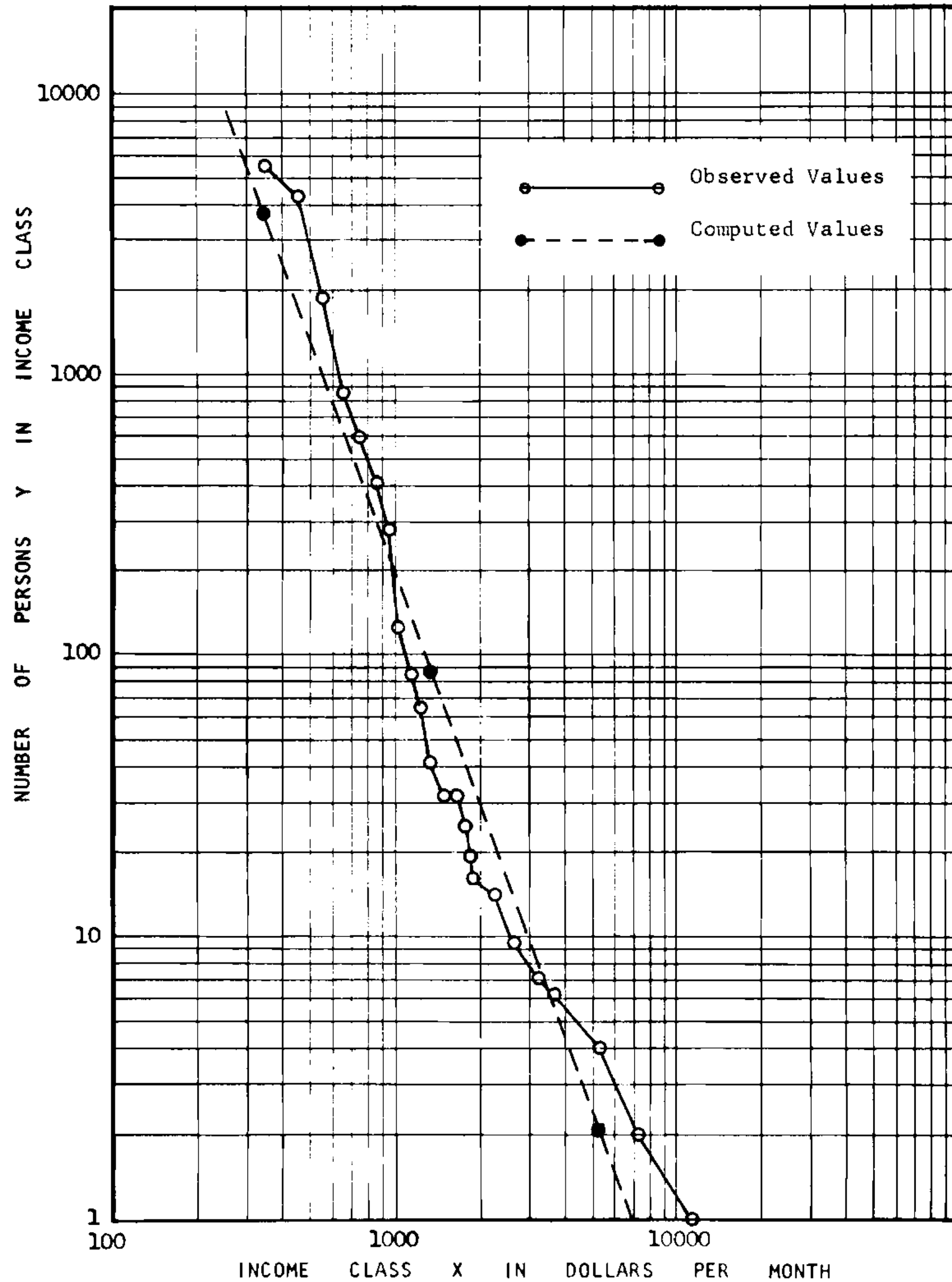

Figure 10. Wage and Salary Distribution of Company I

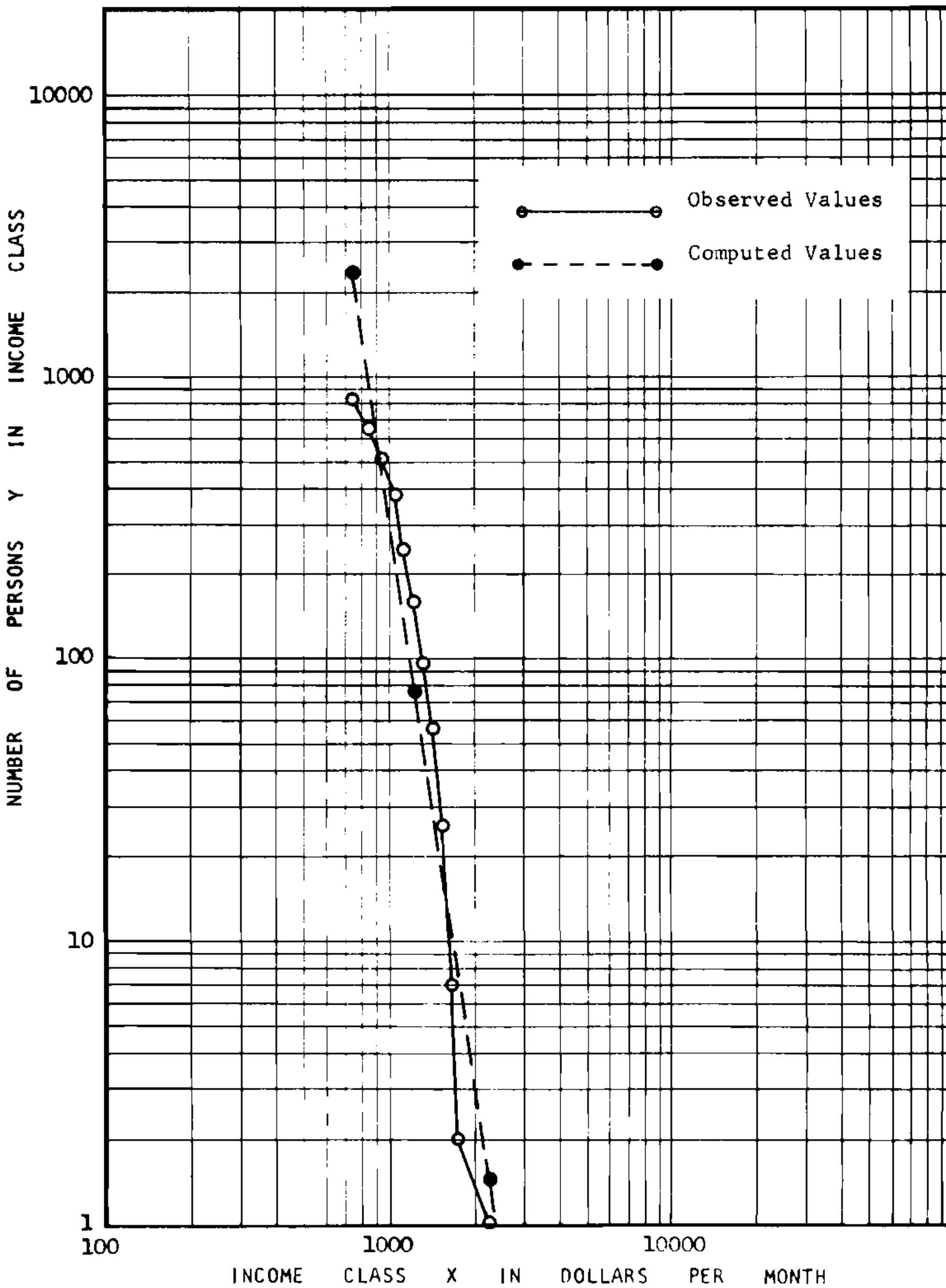

Figure 11. Wage and Salary Distribution of Company J

Table 5. Comparison of the Paretian Constants Among Companies and Industry Groups

COMPANY	ALL VALUES OF INCOME DATA INCLUDED		SEVERAL EXTREME VALUES OF INCOME DATA EXCLUDED	
	Log a	v	Log a	v
A	18.71970	5.75101	27.44517	8.64743
B	17.40295	5.30257	23.19815	7.25197
C	17.54033	5.60786		
D	15.55021	4.75271		
E	12.28348	3.61474	13.47348	4.00349
F	13.10460	4.11735	23.94925	7.78244
G	8.82371	2.06165	13.62872	3.70681
H	16.69499	4.84456	22.75967	6.80503
I	10.24059	2.65419		
J	15.82945	4.54058	27.42231	8.20436
K	15.58194	4.89226	23.09090	7.49069
L	14.28519	4.03003	18.73343	5.45574
M	14.19760	3.79654	17.66811	4.92861
N	11.74137	3.42288		
O	9.53361	2.52642		

INDUSTRY	Log a	v	Log a	v
I	15.99744	4.67811	20.27569	6.06556
II			19.70230	5.72948
III			14.79066	4.05678
IV			19.70974	5.61702

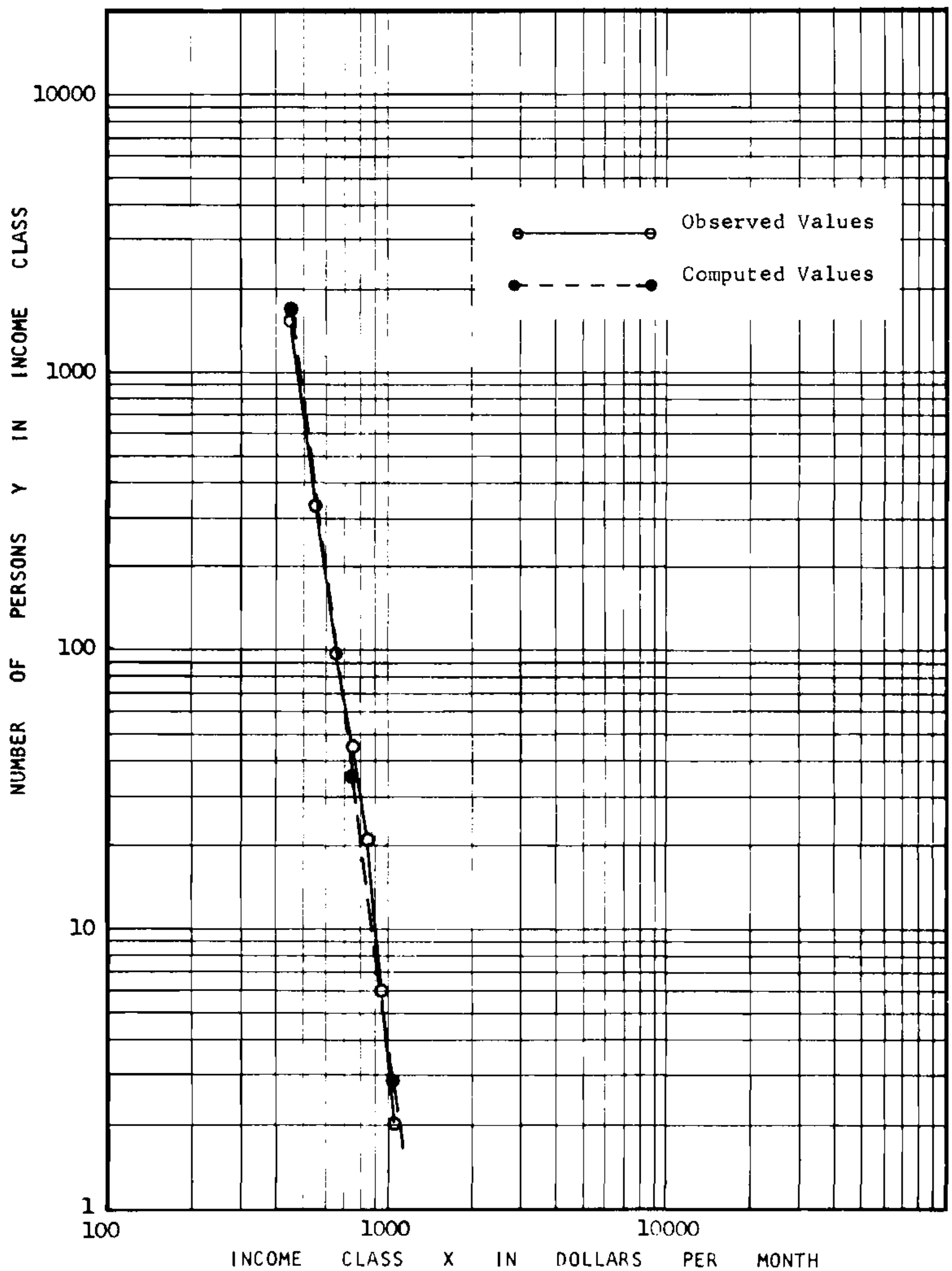

Figure 12. Wage and Salary Distribution of Company K

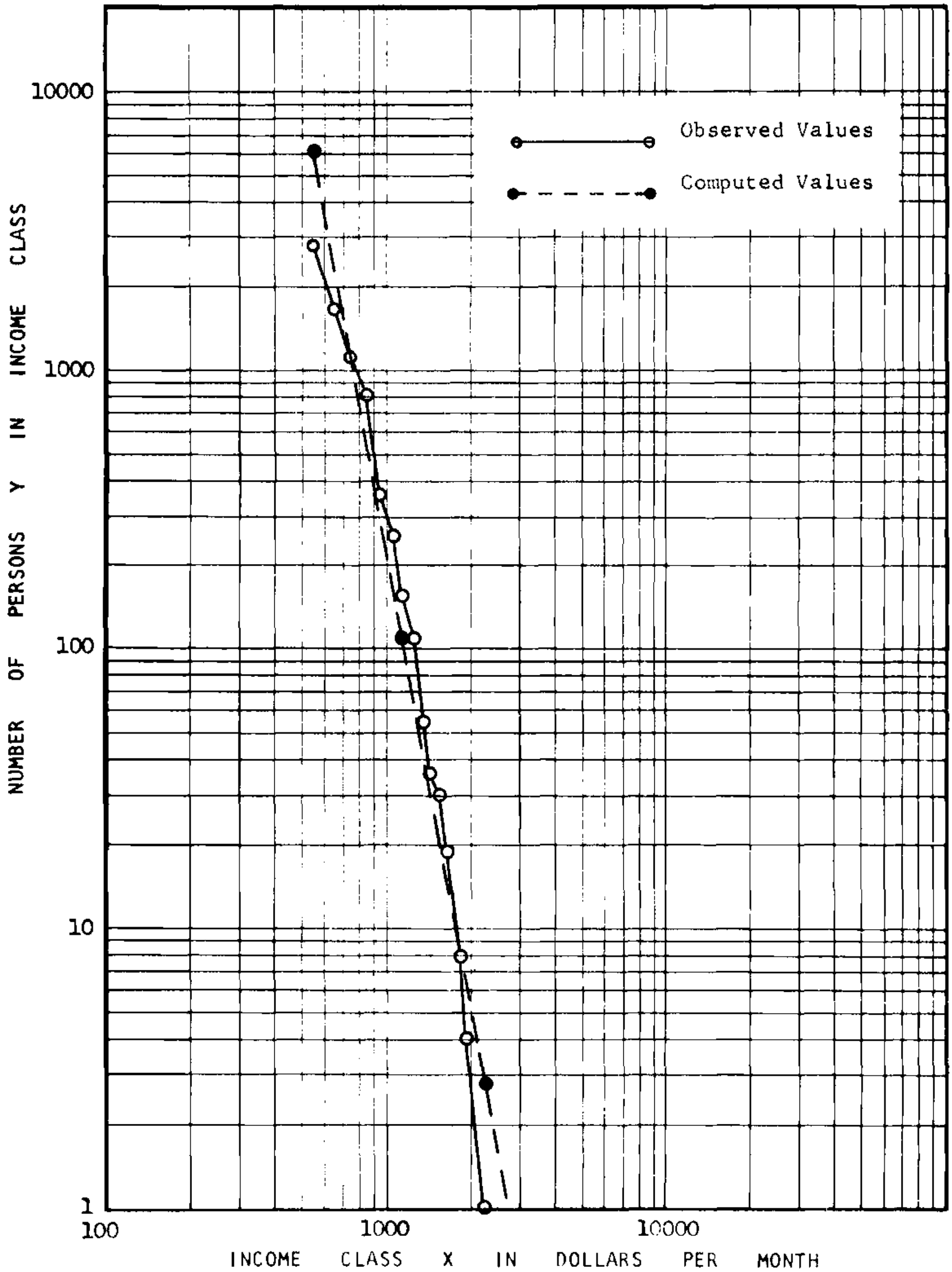

Figure 13. Wage and Salary Distribution of Company L

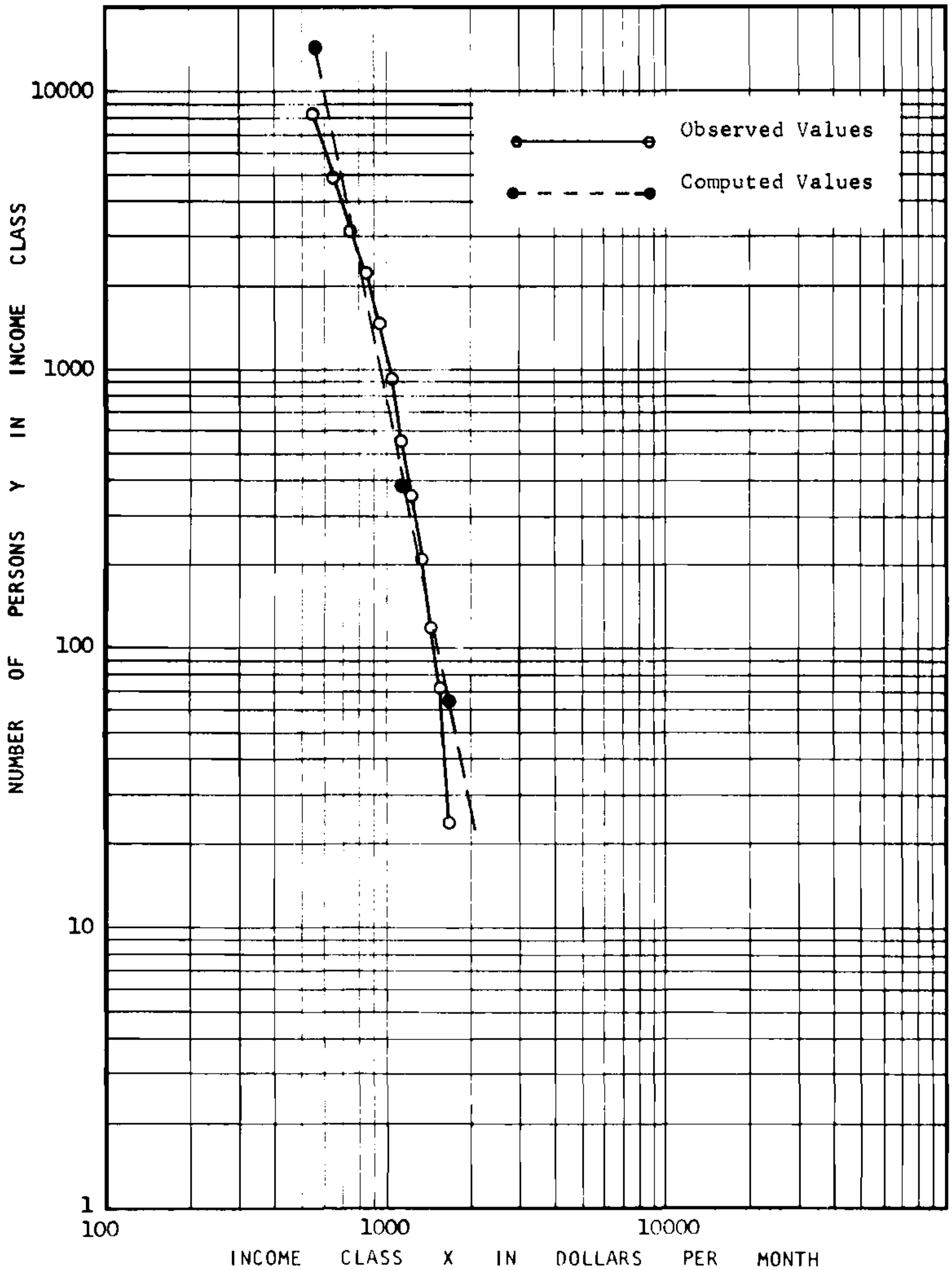

Figure 14. Wage and Salary Distribution of Company M

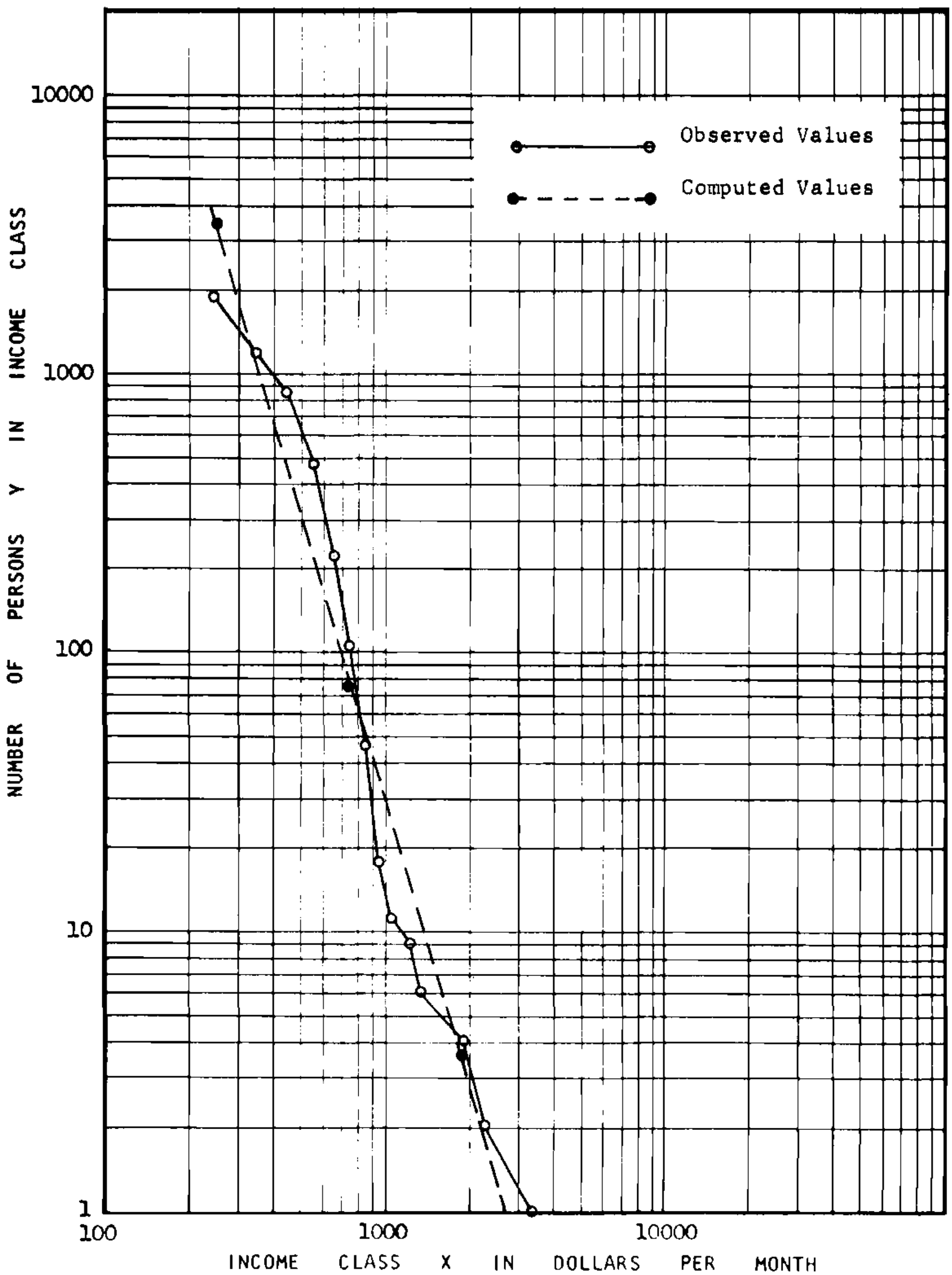

Figure 15. Wage and Salary Distribution of Company N

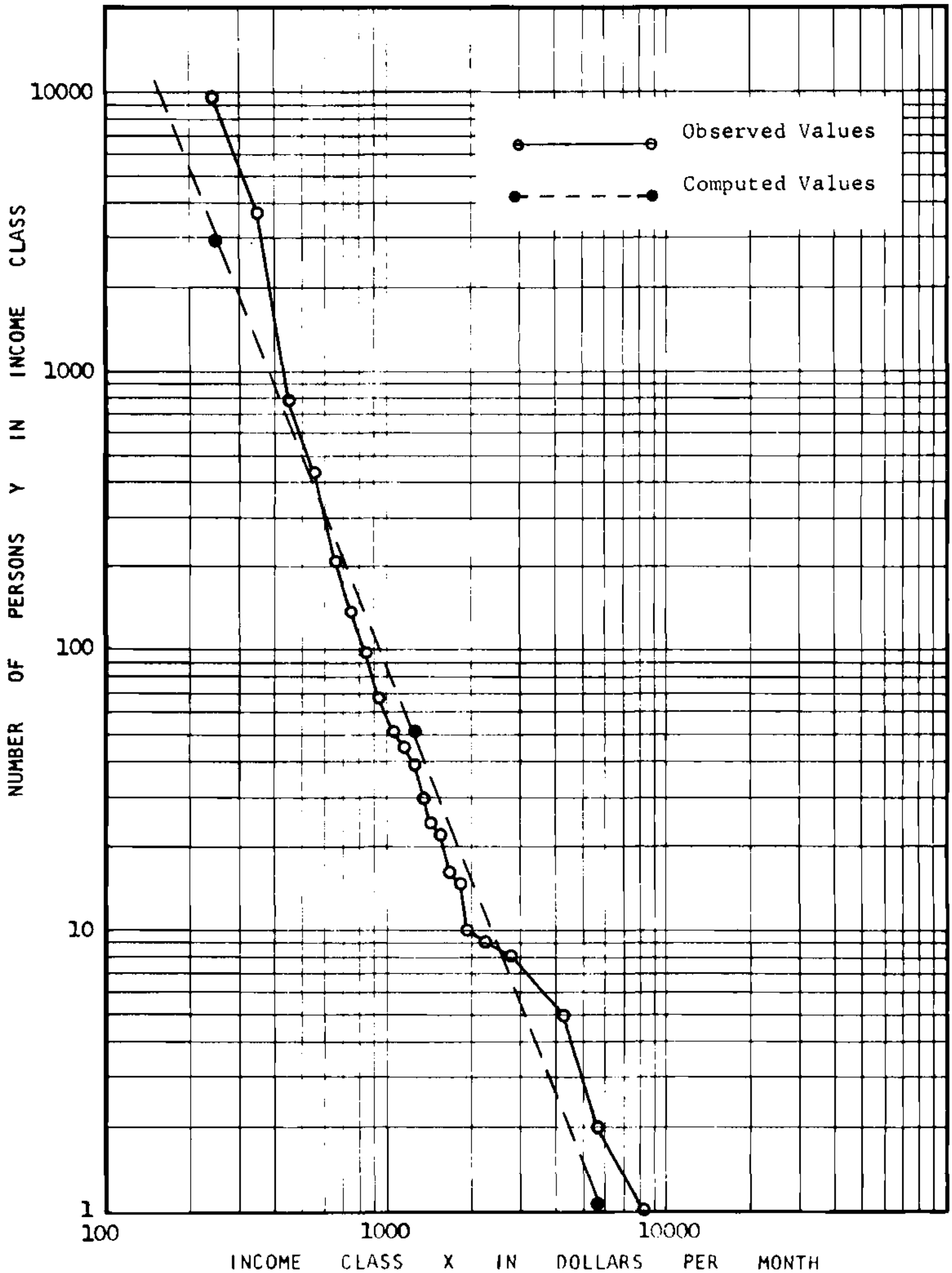

Figure 16. Wage and Salary Distribution of Company O

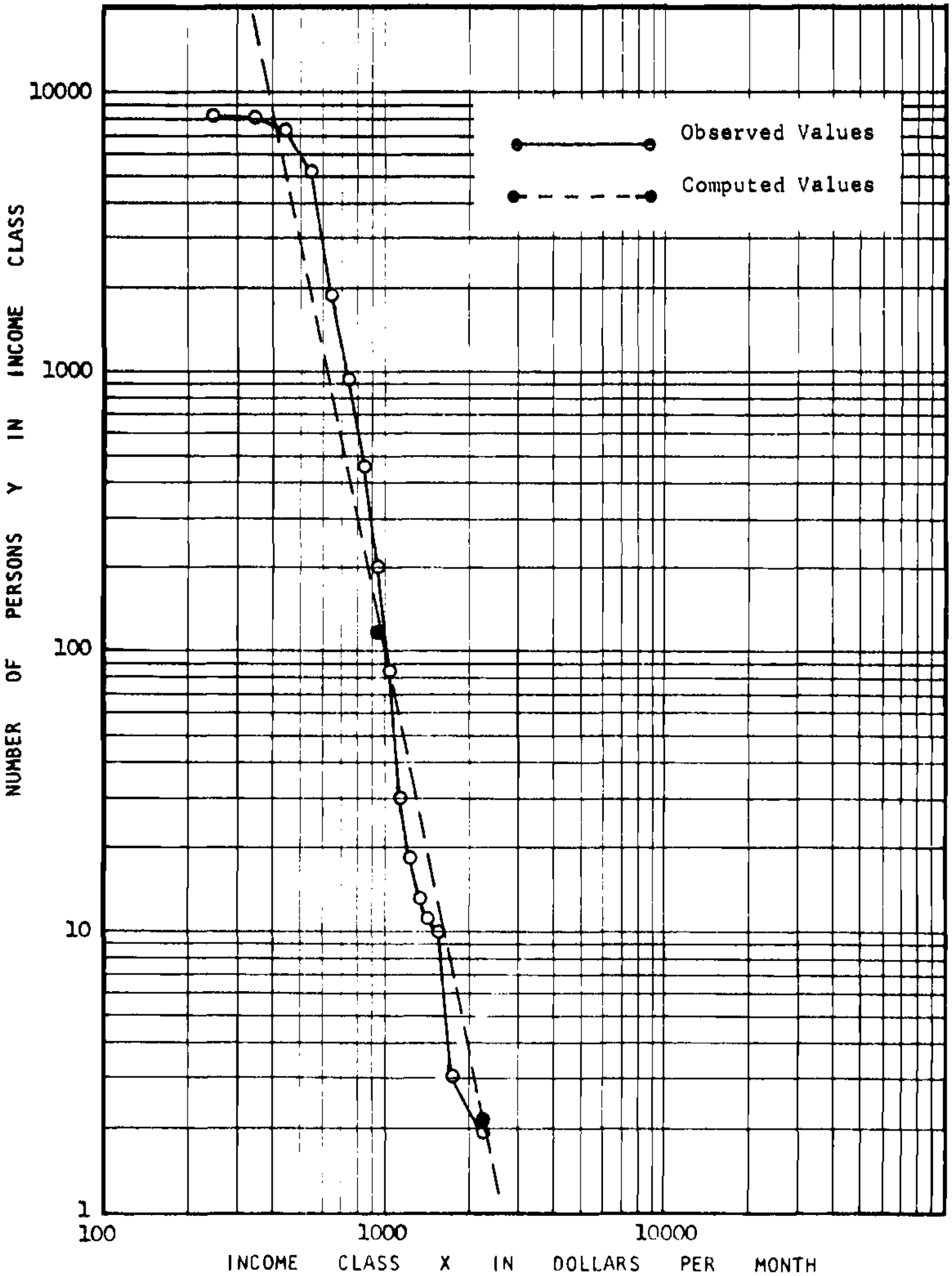

Figure 17. Wage and Salary Distribution of Industry I

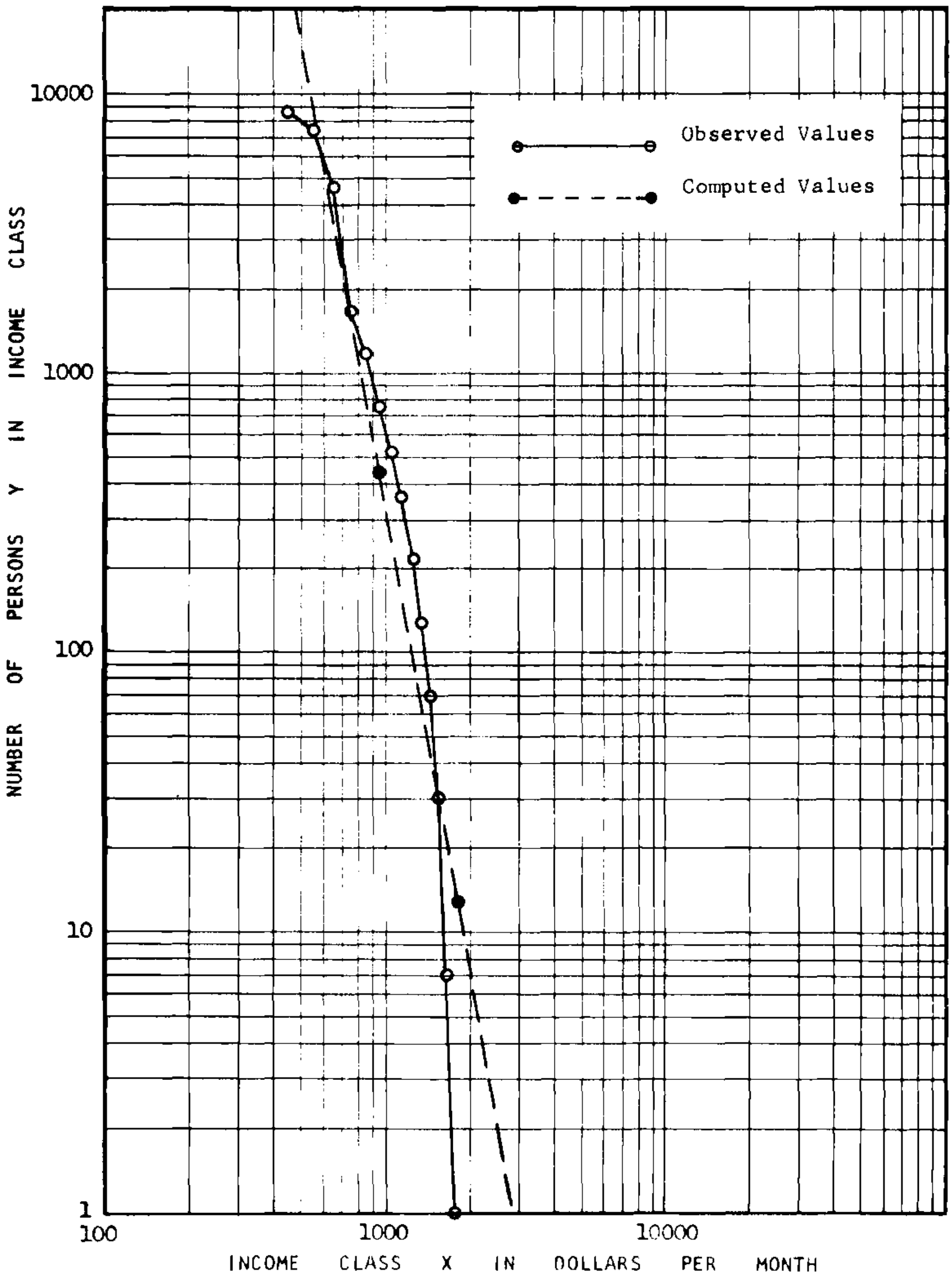

Figure 18. Wage and Salary Distribution of Industry II

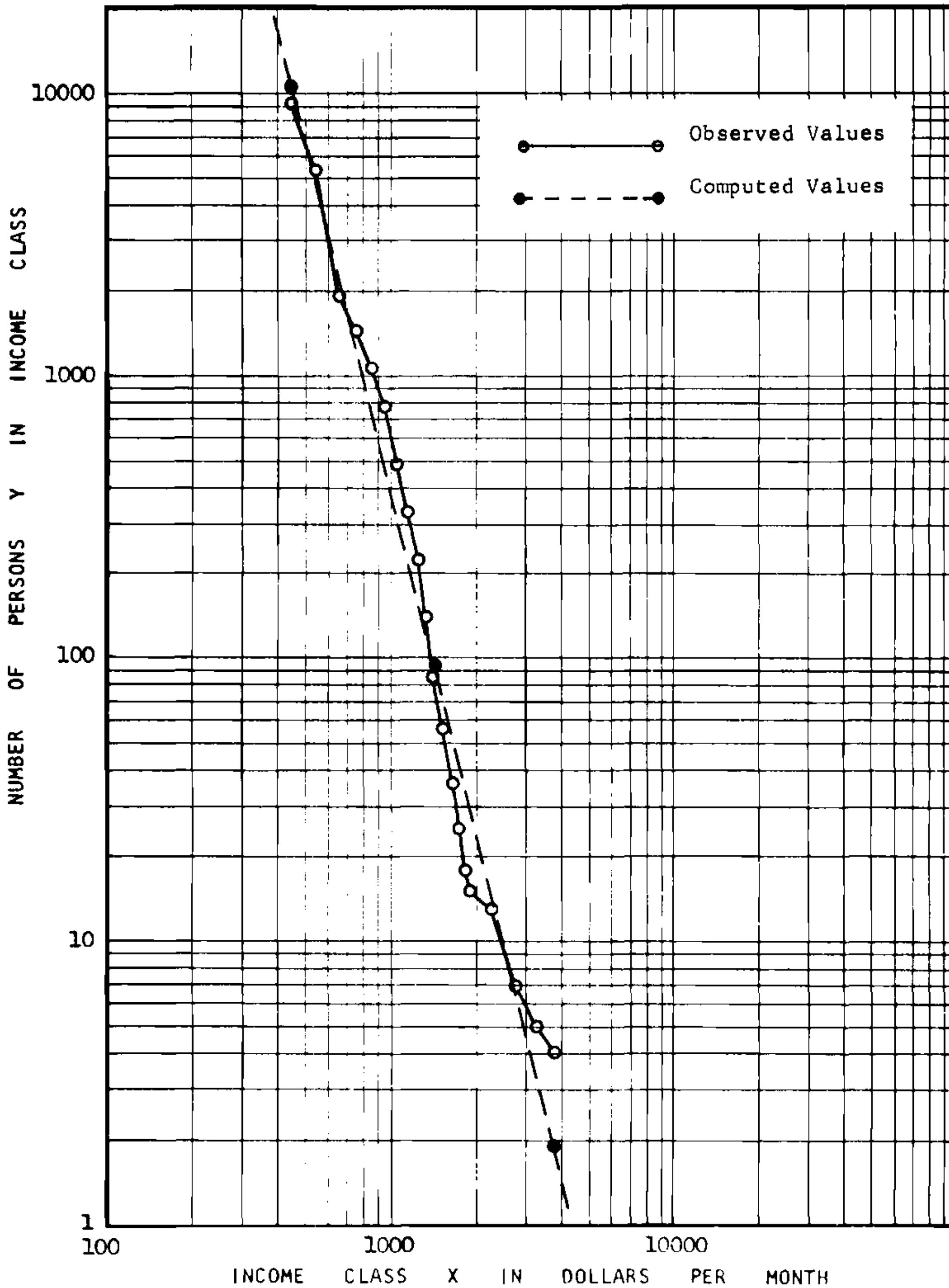

Figure 19. Wage and Salary Distribution of Industry III

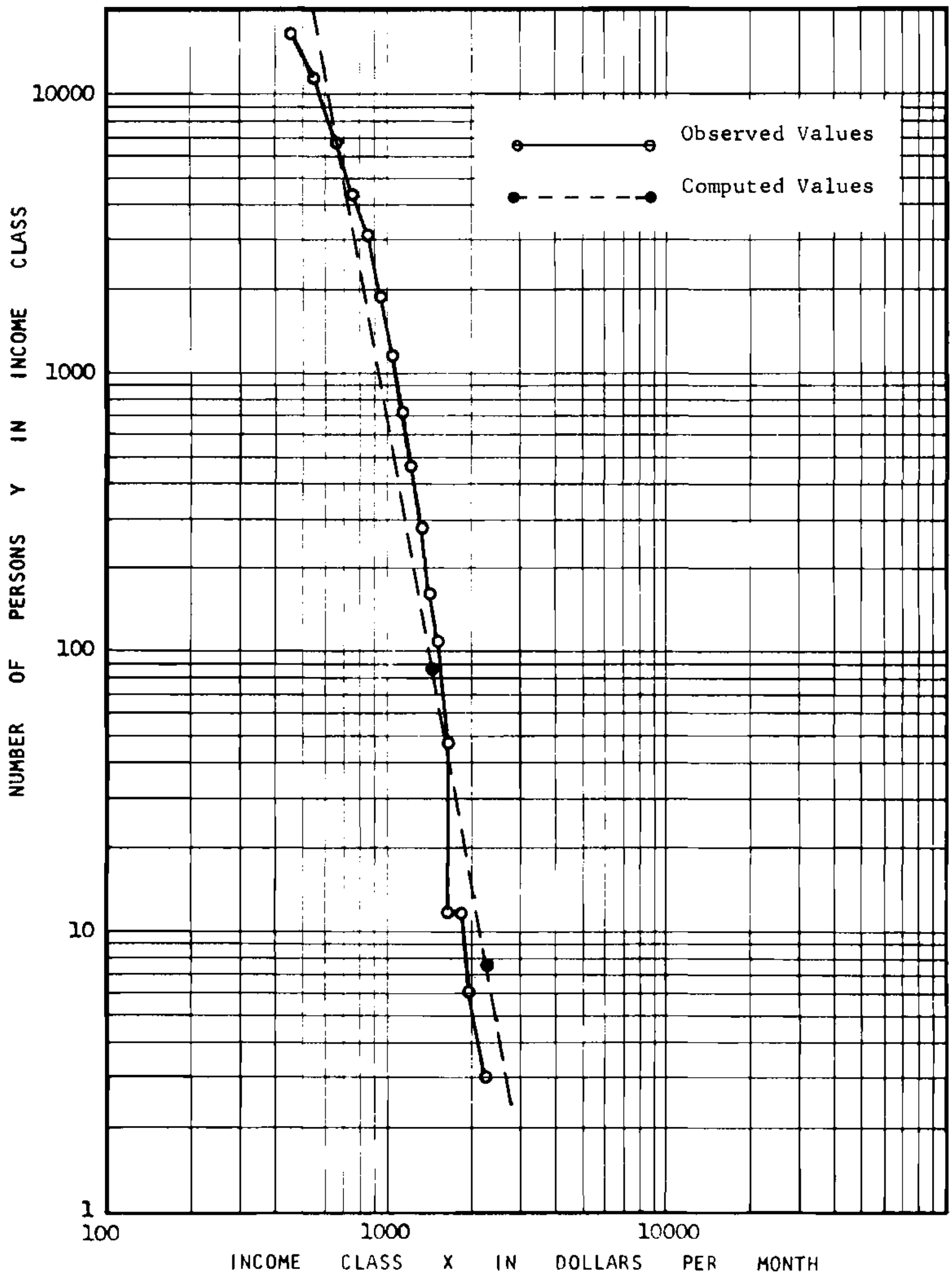

Figure 20. Wage and Salary Distribution of Industry IV

and v in Table 5 is based on computations including all income data. In arriving at the second set of values for these parameters, several extreme values were excluded from income data. In addition to the cumulative frequency tables employed in calculations, several sample calculations illustrating the manner of evaluating a and v by the method of least-squares are included in the Appendix.

The cumulative frequency curves which were obtained for each of the fifteen companies may be classified into one of the following types: (a) those which have a very steep slope; (b) those which have a moderately steep slope; and (c) those which have a small slope.

The above comparisons of slope are relative to one another. The differences in slope used to form the three classifications are for the entire range of the curves which represent all the income data of the particular company. Roughly speaking, the seven companies A, B, C, D, H, J, and K are in group (a), five companies E, F, L, M, and N in group (b), and three companies G, I, and O in group (c). Companies with the largest number of employees fell in groups (b) and (c).

It is of interest to investigate the changes in income distribution by increasing or decreasing the slope of the general curve. Let AB in the accompanying full logarithmic chart, Figure 21, represent the general income curve based on cumulative data and with the assumption that it follows Pareto's general law. Suppose that the income curve between the two incomes X_o and $X_o + X_1$ has been established from the observed income data of the company. Also, let Y_o and $Y_o + Y_1$ indicate the corresponding number of persons in income class X_o and $X_o + X_1$.

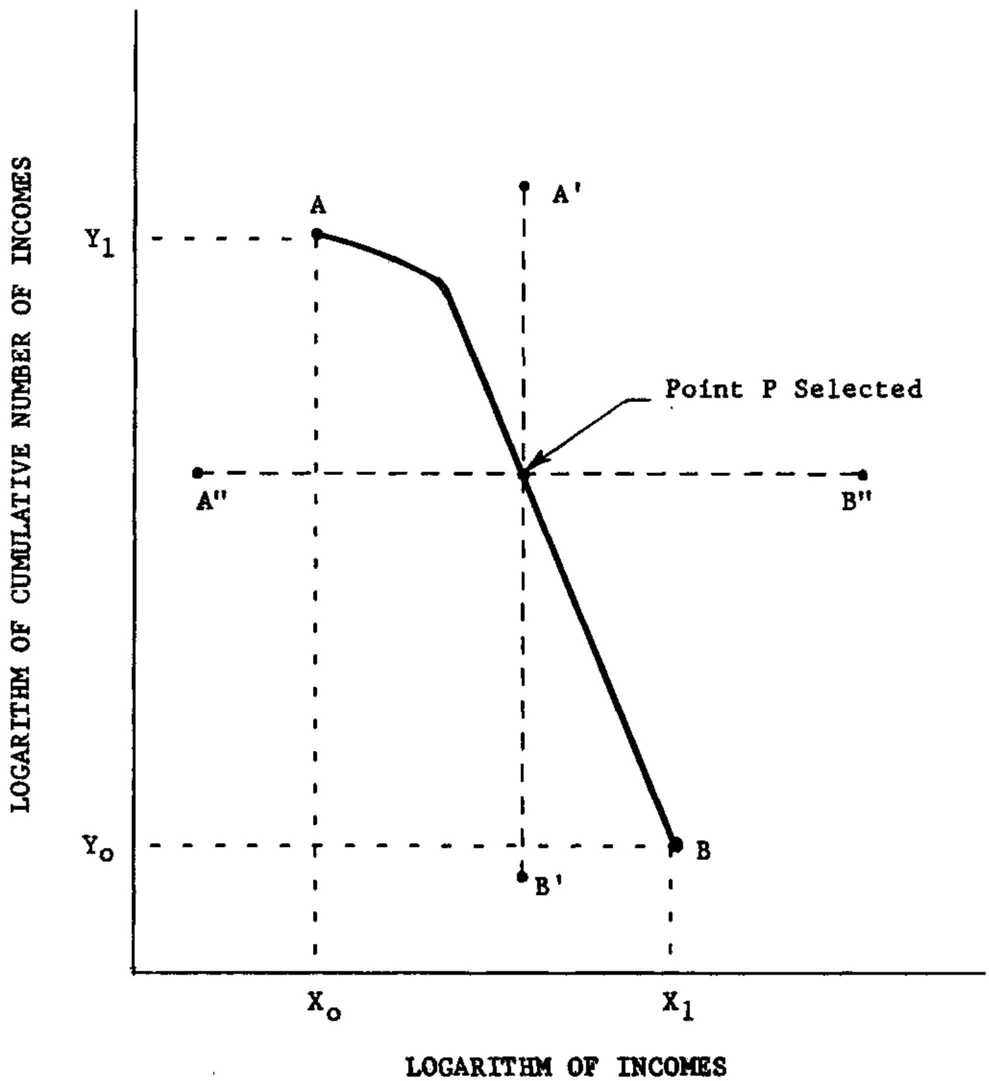

Figure 21. Changing the Slope of the Income Curve

After selecting a point P on the income curve arbitrarily, we let the income curve AB take the position A'B' and A"B" by revolving it about the point P. At position A'B' we observe that all the employees of the company receive the same amount of income. However, at position A"B" we have the same number of people in each income class. Any variation between the two extreme positions would depend on the slope of the curve of each particular company. The facts brought about by the increase or decrease of the slope v might well be interpreted as a sign of increased or decreased inequality of incomes among a group of N employees.

Analyzed data from the 15 companies indicate that the straight-line of slope v (see Figures 2 through 20 and Table 5) on full logarithmic paper should not be extrapolated to lower income levels and in some cases to very high income levels.

One possible explanation as to why the Pareto law does not hold satisfactorily over income range R_l is that the bargaining power of labor unions tends to equalize wage incomes. Similarly, at very high income levels the income tax laws tend to keep the salaries paid to administrative people at a level advantageous to the executive. Thus, the theoretically feasible concept of "infinite income" becomes no longer practical and distortions in tail end of range R_u begin to appear.

If the above proposed explanations are in fact true, our observations would be considered biased and the income data would not naturally lend itself to statistical analysis using empirical laws.

CHAPTER V

SEARCH FOR AN ADEQUATE MATHEMATICAL DESCRIPTION OF INCOME DISTRIBUTION DATA

In research and industry one frequently obtains numerical data relating two or more factors. These factors can be thought of as causes and effects or, in mathematical terms, as independent and dependent variables. The practicing engineer is primarily interested in finding a function that will suitably express the relationships among these variables accurately. The discussion of this chapter will be limited to the dependency between two variables x and y where both are considered to be continuous.

The problem is to determine a specific function, from a given set of numerical data relating two variables, whose curve is consistent with that of the observed data. Knowledge of what is previously known or assumed concerning the relationship between the two variables is important. For the purposes of this study, however, what we desire is some function, called an empirical function, which will express mathematically the relationship between the two variables. Since the exact form of the relationship is unknown, some known relationship that will express the data will be determined for the range of interest of the variables.

There are two basic techniques for determining which family of curves will most accurately express the data in question. These methods, known as graphical and finite differences, rely primarily on the principle that for a large number of relationships certain constant incremental changes in x result in certain constant changes in y. The former method utilizes this principle through the procedure of an appropriate graphical plot while the latter technique is applied by mathematical manipulation.

Once a suitable form for a particular set of observed data has been determined, the next procedure is to calculate the unknown coefficients. Of the different methods available the method of least-squares and selected-points are the most widely used. The former incorporates a principle whereby the coefficients of the general curve are determined as a result of minimizing the sum of the squares of the critical distances between the observed points and the plotted empirical function. The method of selected-points involves selecting a certain number of widely-spaced points of the smoothed data, usually equivalent in number to the number of unknown coefficients, and solving the resultant simultaneous equations. For a detailed discussion of the above methods the reader is referred to references (22), (23), and (24) in the bibliography.

One usually has two distinct purposes in mind, when obtaining a mathematical expression to represent the relation between the variables. These purposes are: (a) to determine the physical law underlying the observed quantities; and (b) to obtain a simple formula, which may or may not have a physical basis, but by which an approximate value of one variable may be computed from a given value of the other variable. In the first case correctness of form is a necessary consideration. In the second, correctness of form is generally considered secondary to simplicity and convenience. It is with the latter of these that the author is mostly concerned.

To describe the entire range of incomes by an empirical function or functions, three distinct approaches were undertaken. For sake of simplicity, the three different methods of attack shall be referred to as Method I, Method II, and Method III. Method I is to formulate a new function from the given data such that this function would describe the relation between y and x adequately and for the entire range of incomes.

Method II is to modify the original Pareto formula by introducing a function $\Phi(x)$. Since Pareto's law holds reasonably good for values of x greater than or equal to a certain income, data falling in range R_u can be described using the expression $y = ax^{-v}$. Data falling outside this range was categorized to be in the lower income range R_1 of the distribution curve expressed by the modified equation $y = ax^{-v}\Phi(x)$. Selection of the function $\Phi(x)$ is arbitrary so long as it meets the desirable requirements for each particular establishment under investigation.

Method III is similar to Method II, except that the lower salary range (R_1) of the general curve was described by the function $y = Ne^{-c \cdot (x-x_0)^2}$, where y = number of persons in income class x; n = total number of employees in the company; x_0 = lowest income class interval for the company; x = income class; and c = constant to be determined from data of the company.

Method I

Income data of company E was tabulated as shown in Table 6. Table 20 in the Appendix is a cumulative frequency tabulation of the data for the same company. The values for x versus values for y, and values for x versus cumulative values for y are plotted on ordinary graphical paper (refer to Figure 22 of this chapter). The resulting two curves were compared with curves having approximately the same shape, and the appropriate function describing these curves was selected. The next step was to determine the values of the parameters in the selected function from the observed data and using the techniques discussed earlier in the chapter.

The general equation selected is of the form

$$y = ax^{b}e^{cx} \tag{13}$$

where, y = number of persons in income class x; x = income class, and a, b, and c are the three parameters to be determined from data of the company. Based on the method of finite differences and using tabulated data in Table 6, calculations on page 94 of the Appendix were made to determine the values of the parameters.

From the computations a, b, and c were found to be 1.13, -0.264, and -0.0031, respectively. Substituting these values in equation (13) above, we have

$$y = 1.13\ x^{-0.264}e^{-0.0031\ x}. \tag{14}$$

Table 6. Tabulated Income Data of Company E

y	x	lg x	lg y	Δ lg x	Δ lg y	Δ_1 lg y
3	$ 250	2.397	0.477	0.147	2.040	+1.176
329	350	2.544	2.517	0.109	0.350	-1.819
737	450	2.653	2.867	0.087	-0.627	-2.566
174	550	2.740	2.240	0.072	-0.516	-2.240
53	650	2.812	1.724	0.063	-0.071	-1.724
45	750	2.875	1.653	0.054	-0.256	-1.653
25	850	2.929	1.397	0.048	-0.318	
12	950	2.977	1.079	0.044	-0.381	
5	1,050	3.021	0.698	0.039	-0.397	
2	1,150	3,060	0.301	0.036	-0.301	
1	1,250	3.096		0.034	+0.301	
2	1,350	3.130	0.301	0.031	-0.301	
1	1,450	3.161		0.029	+0.778	
6	1,550	3.191	0.778	0.027	-0.778	
0	1,650	3.217		0.026		
1	1,750	3.243		0.024		
0	1,850	3.267		0.023		
0	1,950	3.290				
1	2,250	3.352				
Totals	20,950	47.186	17.031			

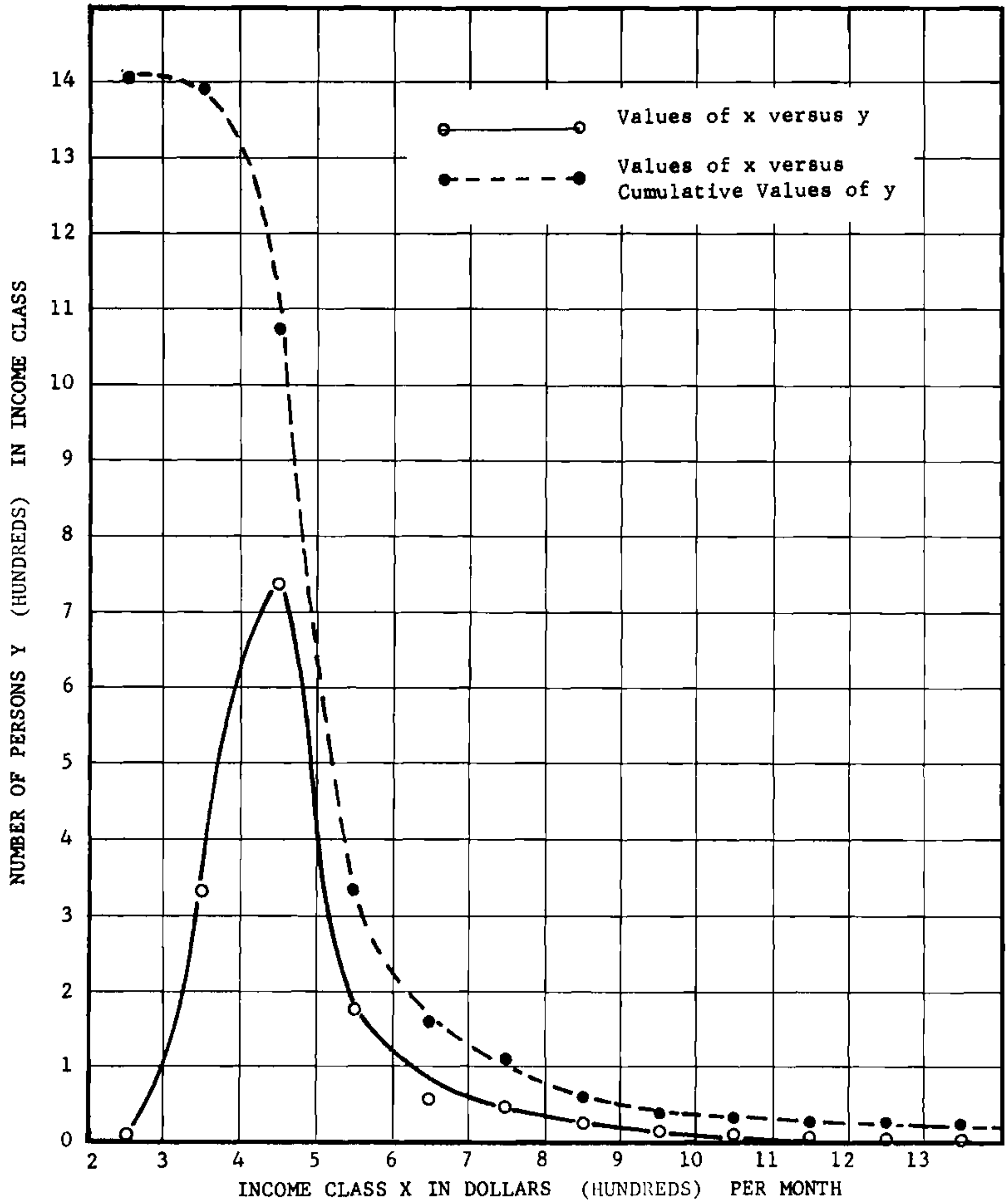

Figure 22. Curve Fitting to Income Data of Company E

The computed values of y using equation (14) were far below the observed values over the entire range of the income curve.

Taking income data corresponding to the curve for x versus cumulative y, the general equation (13) was tested again. Results of the calculations based on the <u>selected-points</u> method led to the following equation:

$$y = 5.17\, x^{-1.64} e^{-0.036\, x}. \tag{15}$$

The calculations are on page 95 in the Appendix. Again the computed values of y using relationship (15) were significantly different from the observed values. Therefore, the general equation (13) was abandoned.

Method II

For the purposes of modifying the adopted Pareto equation only over the range R_1 as shown in figure 23 of this chapter, the following exponential and hyperbolic functions were utilized:

$$\Phi(x) = (1-e^{-cx}); \text{ and } y = ax^{-v}(1-e^{-cx}), \tag{16}$$

$$\Phi(x) = [1-e^{-(cx)^2}]; \text{ and } y = ax^{-v}\,[1-e^{-(cx)^2}], \tag{17}$$

$$\Phi(x) = \tanh(cx); \text{ and } y = ax^{-v}\,[\tanh(cx)], \tag{18}$$

$$\Phi(x) = \sinh(cx); \text{ and } y = ax^{-v}\,[\sinh(cx)]. \tag{19}$$

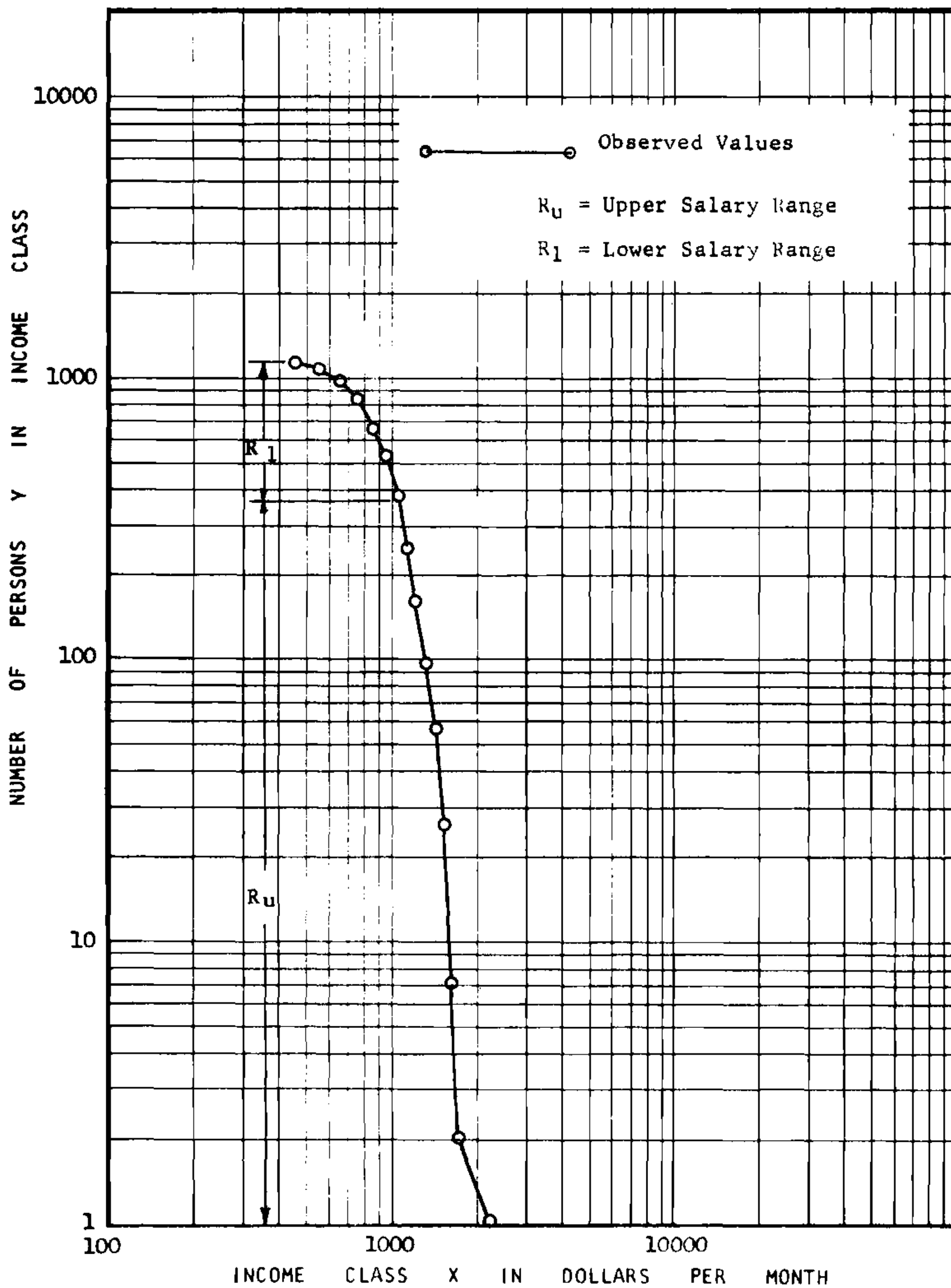

Figure 23. Separation of Income Curve into Upper and Lower Ranges

It should be noted that the selected exponential and hyperbolic functions in equations (16), (17), (18), and (19) have one common characteristic. For values of $x = 0$, $\Phi(x) = 0$, and as x approaches infinity $\Phi(x) = 1$.

The expressions (16), (17), (18), and (19) were applied to income data of company J. A comparative summary of modified Pareto values versus the observed values of income data in range R_1 is presented in Table 7. It is observed from Table 7 that the computed and the observed values differ significantly. This leads us to the conclusion that the task of modifying Pareto distribution in range R_1 is not such an easy one.

Method III

We begin with the assumption that the upper salary range R_u is described adequately by the Pareto expression $y = ax^{-v}$. Now, the problem is to establish a new independent function such that it describes all data over the range R_1. The cumulative income data of Company J was plotted on ordinary graph paper (Figure 24) and segmented into income ranges R_u and R_1.

Upon close observation of income data for various companies and in range R_1, the formula below was established based on the general shape of income curves.

$$y = Ne^{-c(x-x_0)^2} \tag{20}$$

In equation (20) the only unknown parameter is c. The value of c will

Table 7. Comparison of Modified Pareto Values Versus Actual Values in Range R_1 (Company J)

Income Class Mark (x)	$y=ax^{-v}(1-e^{-cx})$	$y=ax^{-v}[1-e^{-(cx)^2}]$	$y=ax^{-v}\tanh(cx)$	$y=ax^{-v}\sinh(cx)$	y Actual
$ 450					
550	19,675	7,292	22,296	23,521	1,126
650	6,459	2,785	7,427	7,999	990
750	2,537	1,269	2,953	3,257	828
850	879	504	1,034	1,171	651
950	435	281	515	601	508
1,050	187	132	223	269	377

NOTE: The parameter c was computed at x = \$550 and x = \$1,050 and by taking the average corresponding to these values of x.

be determined using the inflection-point concept of calculus. After taking the first and second derivative of formula (20), we set it equal to zero and solve for x at the inflection-point. Substituting this value of x in expression (20), we compute the corresponding value of y at the inflection-point.

We have $y = Ne^{-c(x-x_0)^2}$ from (20), and

$$\frac{dy}{dx} = -\ 2cN(x-x_0)e^{-c(x-x_0)^2}. \tag{21}$$

Then, the second derivative of (20) becomes

$$\frac{d^2y}{d^2x} = 2cNe^{-c(x-x_0)^2}\ [2c(x-x_0)^2-1]. \tag{22}$$

Setting $\frac{d^2y}{d^2x} = 0$, and solving for the inflection-point value of x we get,

$$x = x_0 + \frac{1}{\sqrt{2c}}. \tag{23}$$

By substituting $(x_0 + \frac{1}{\sqrt{2c}})$ for x in equation (20),

$$y = \frac{N}{\sqrt{e}} = 0.606\ N. \tag{24}$$

Thus, for all companies under investigation the inflection points x and y for the adopted general equation over income range R_1 is given by equations (23) and (24). Equation (24) states that, in round figures, R_1 comprises about 40% of all company's employees, while R_u comprises

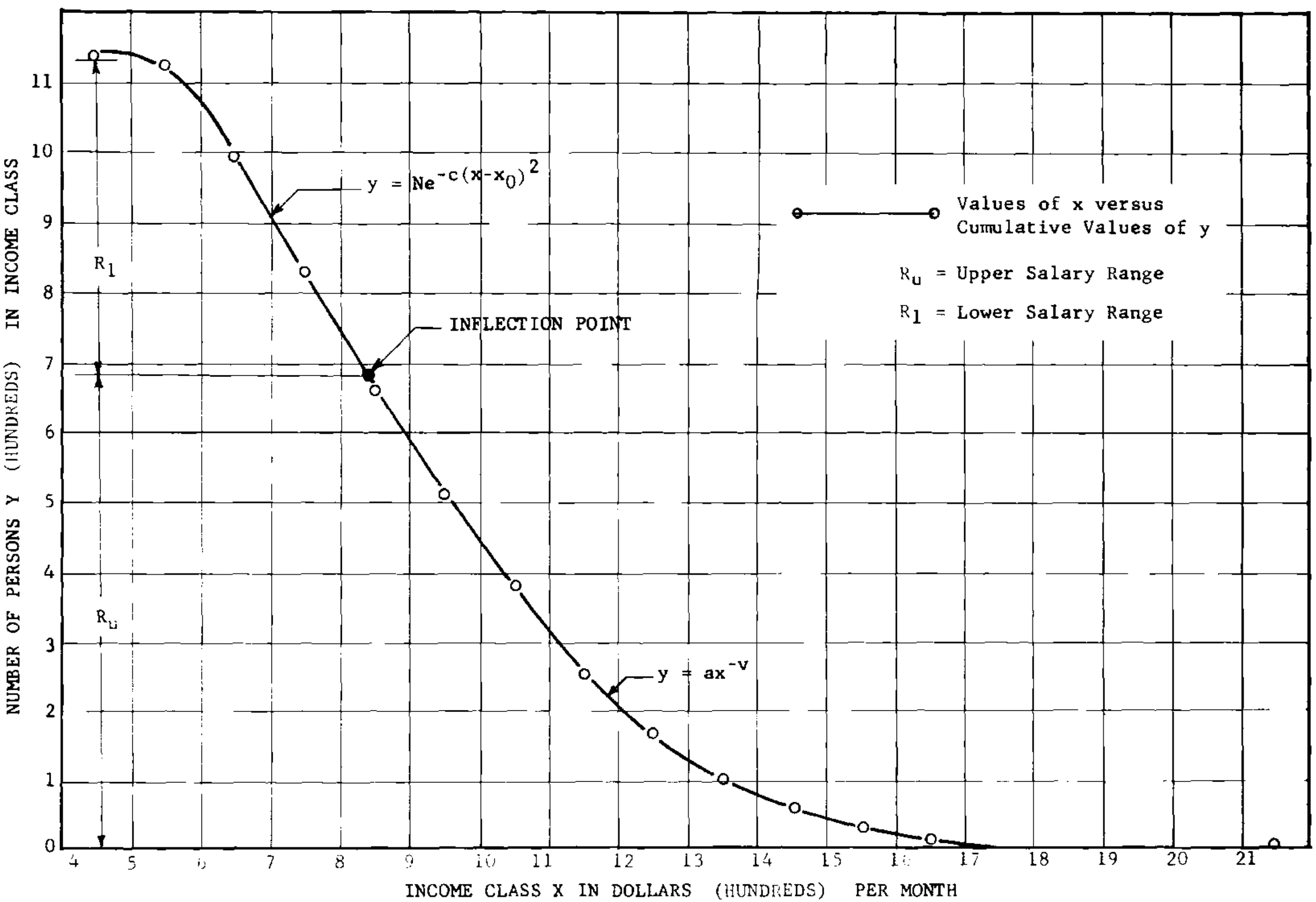

Figure 24. Distribution of Incomes for Data of Company J

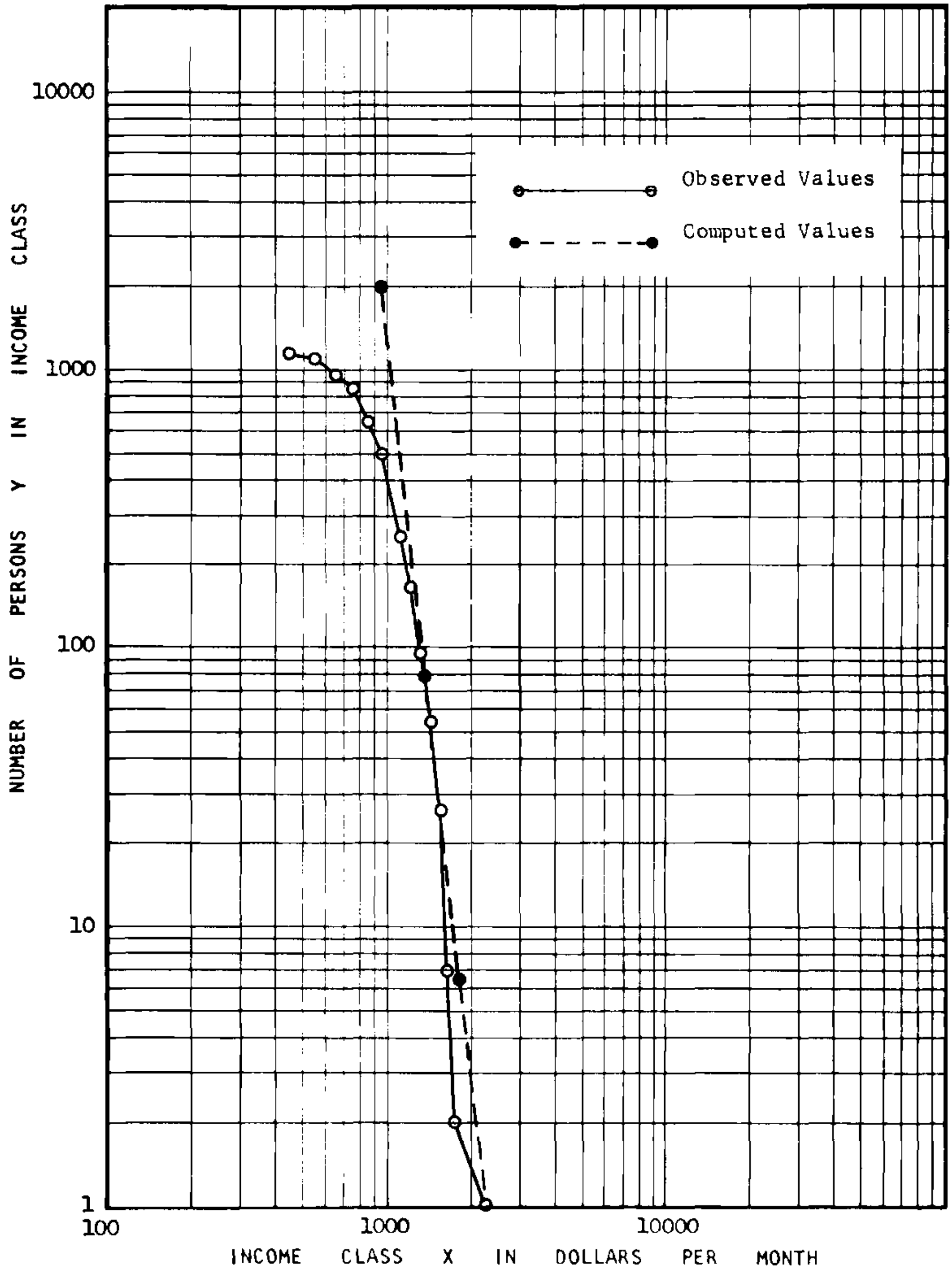

Figure 25. Pareto Curve Fitted to Data of Company J in Range R_u

Table 8. Cumulative Frequency Tabulation for Data of Company J

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 450	1128							
550	1126							
650	990							
750	828							
850	651							
950	508	2.97772	2.07586	8.05729	8.86682	24.43029	2.99202	981.79
1,050	377	3.02119	2.57634	7.78361	9.12759	24.78693	2.63538	431.90
1,150	250	3.06070	2.39794	7.33938	9.36788	25.11108	2.31123	204.76
1,250	162	3.09691	2.20952	6.84269	9.59085	25.40816	2.01415	103.31
1,350	97	3.13033	1.98677	1.21925	9.79897	25.68235	1.73996	54.95
1,450	56	3.16137	1.74819	5.52668	9.99426	25.93702	1.48529	30.57
1,550	26	3.19033	1.41497	4.51422	10.17821	26.17462	1.24769	17.69
1,650	7	3.21748	0.84510	2.71909	10.35218	26.39736	1.02495	10.59
Totals		31.45125	16.18572	49.97847	99.03118			

NOTE: Values corresponding to the $1,750 and $2,250 class mark are not shown here.

60% of the total.

Since the total number of employees N in the company is known and e = 2.71828, the value of y can be determined from expression (24). Using the smoothed cumulative frequency curves (for example, Figure 26) for income data of a company, the corresponding value of x is established. Then we substitute this value of x in expression (23) and solve for the unknown parameter c. Solving for c, we have

$$c = \frac{1}{2(x-x_0)^2} \, . \tag{25}$$

Having determined c we go back to equation (20) and apply it to data of the company in question. Experimental computations indicate that dividing the value of c by two before substituting in equation (20) appreciably improves the overall fit of the income data. For the purposes of this study the value c or $\frac{c}{2}$ will be used in the general formula depending upon which gives better results.

Application of Method III will be illustrated using data of Company J. From the cumulative frequency Table 8, N = 1128 and x_0 = $450. By equation (24)

$$y = \frac{N}{\sqrt{e}} = \frac{1128}{1.649} = 683.63$$

The corresponding value of x from the curve in Figure 24 is found to be $840 approximately. Then we have

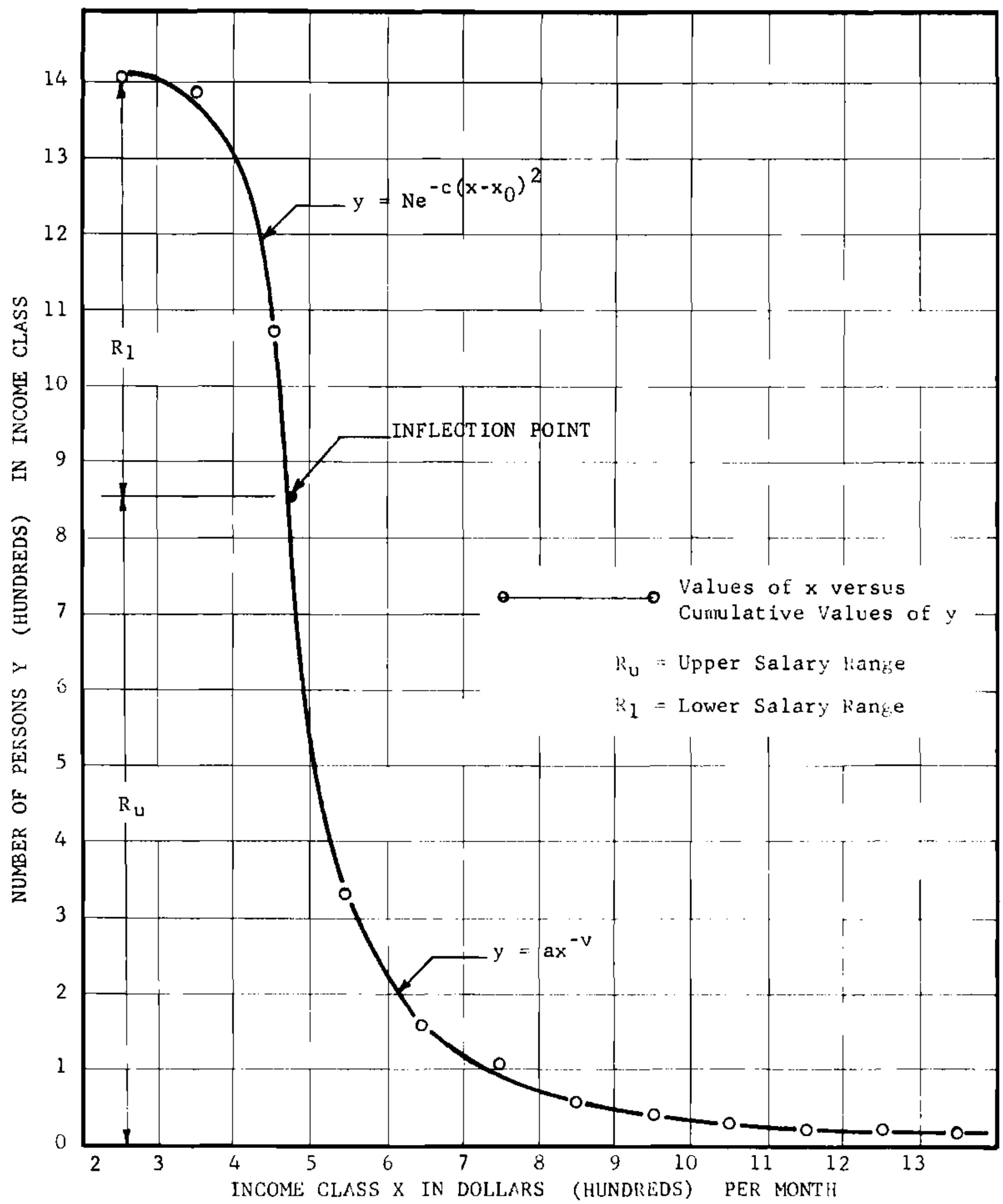

Figure 26. Distribution of Incomes for Data of Company E

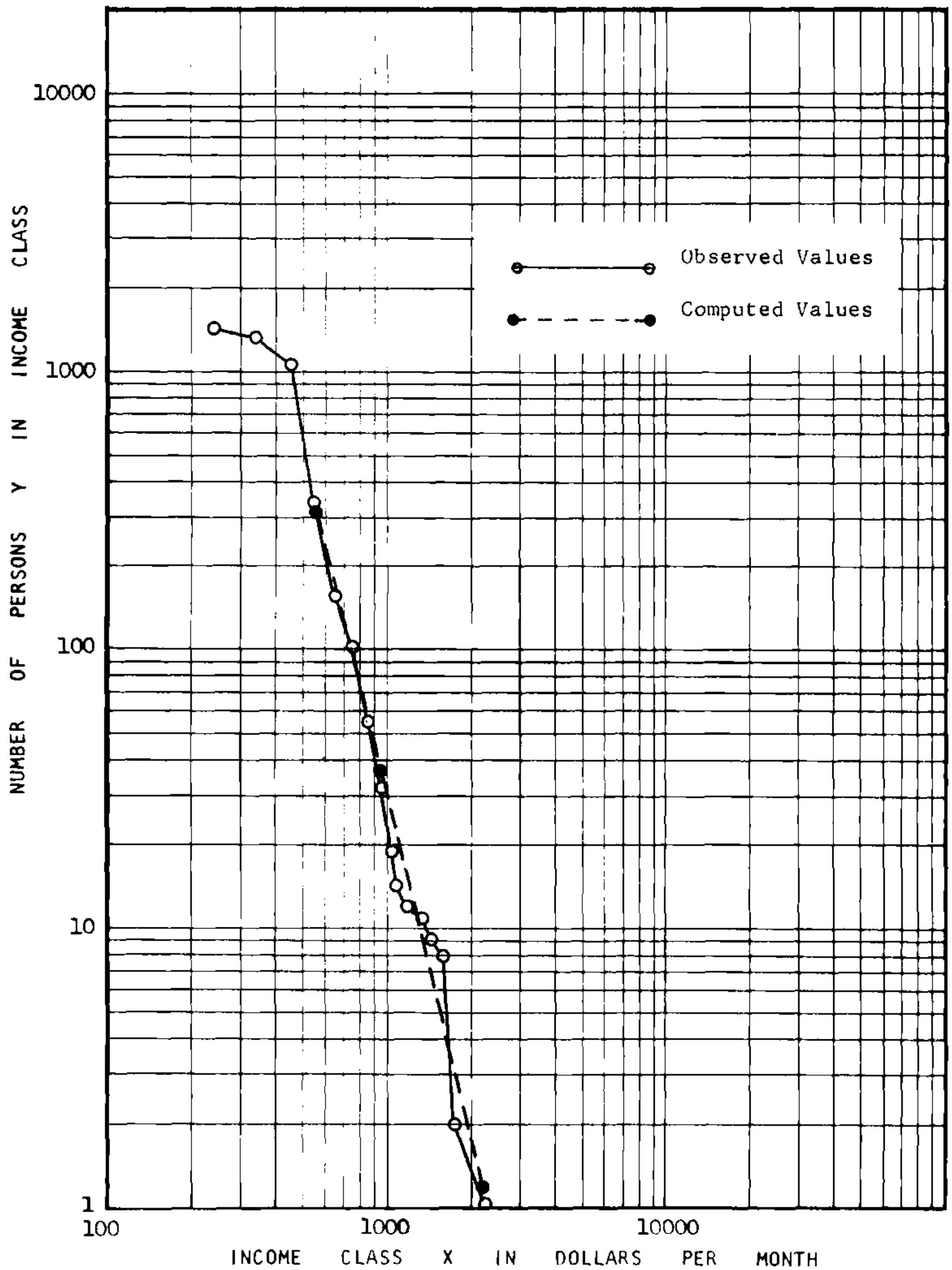

Figure 27. Pareto Curve Fitted to Data of Company E in Range R_u

Table 9. Cumulative Frequency Tabulation for Data of Company E

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	1,397							
350	1,394							
450	1,065							
550	328	2.74036	2.51587	6.89439	7.50957	10.97100	2.50248	318.04
650	154	2.81291	2.18752	6.15330	7.91246	11.26146	2.21202	162.94
750	101	2.87506	2.00432	5.76254	8.26597	11.51027	1.96321	91.88
850	56	2.92942	1.74819	5.12118	8.58150	11.72790	1.74558	55.67
950	31	2.97772	1.49136	4.44085	8.86682	11.92127	1.55221	35.66
1,050	19	3.02119	1.27875	3.86335	9.12759	12.09530	1.37818	23.88
1,150	14	3.06070	1.14613	3.50796	9.36788	12.25348	1.22000	16.59
1,250	12	3.09691	1.07918	3.34212	9.59085	12.39845	1.07503	11.89
1,350	11	3.13033	1.04139	3.25989	9.79897	12.53224	0.94124	8.73
1,450	9	3.16137	0.95424	3.01671	9.99426	12.65651	0.81697	6.56
1,550	8	3.19033	0.90309	2.88116	10.17821	12.77245	0.70103	5.02

Table 9. Cumulative Frequency Tabulation for Data of Company E (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$1,750	2	3.24304	0.30103	0.97625	10.51731	12.98347	0.49001	3.09
2,250	1	3.35218			11.23711	13.42042	0.05306	1.13
Totals		39.59152	16.65107	49.21970	120.94850			

Table 10. Cumulative Frequency Tabulation for Data of Company F

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	139							
450	137							
550	118							
650	89	2.81291	1.94939	5.48346	7.91246	21.89130	2.05795	114.27
750	42	2.87506	1.62325	4.66694	8.26597	22.37498	1.57427	37.52

$$c = \frac{1}{2(x-x_0)^2} = \frac{1}{2(840-450)^2}$$

$$c = \frac{1}{304200} = 3.289 \times 10^{-6}.$$

Thus for Company J income data in range R_1 is described by the expression

$$y = 1128\ [e^{-3.289 \times 10^{-6}(x-x_0)^2}]. \tag{26}$$

$$\text{When } x = \$450,\ y = 1128\ e^0 = 1128.00$$

$$x = \$550,\ y = 1128\ e^{-0.0328}$$
$$= 1128\ (0.971) = 1095.29$$

$$x = \$650,\ y = 1128\ e^{-0.1312}$$
$$= 1128\ (0.878) = 990.38$$

$$x = \$750,\ y = 1128\ e^{-0.2952}$$
$$= 1128\ (0.741) = 835.85$$

$$\text{and at } x = \$850,\ y = 1128\ e^{-0.5248}$$
$$= 1128\ (0.589) = 564.39$$

In the manner described above Method III was applied to income data of companies E, F, and Industry Group I. The reader is referred to

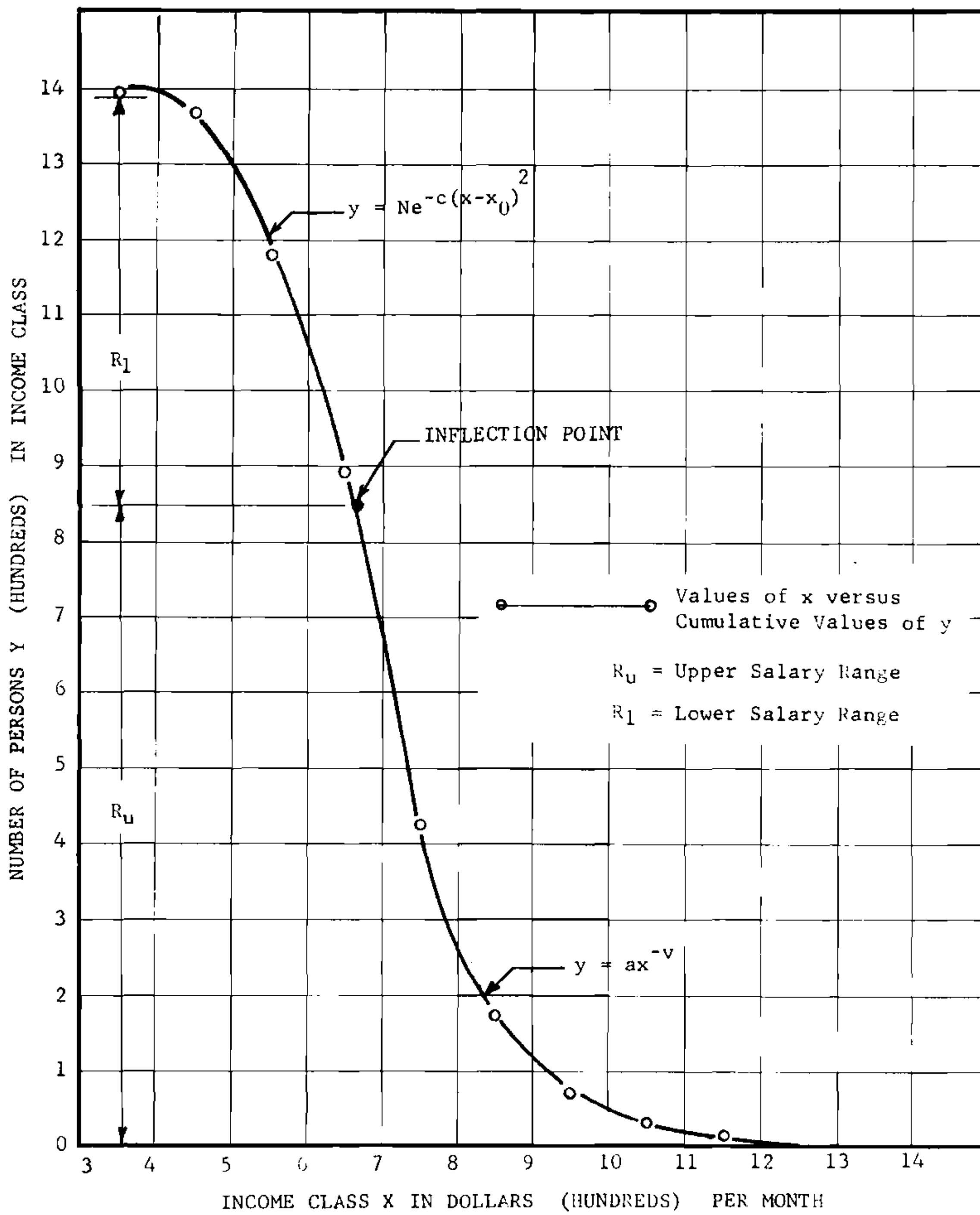

Figure 28. Distribution of Incomes for Data of Company F

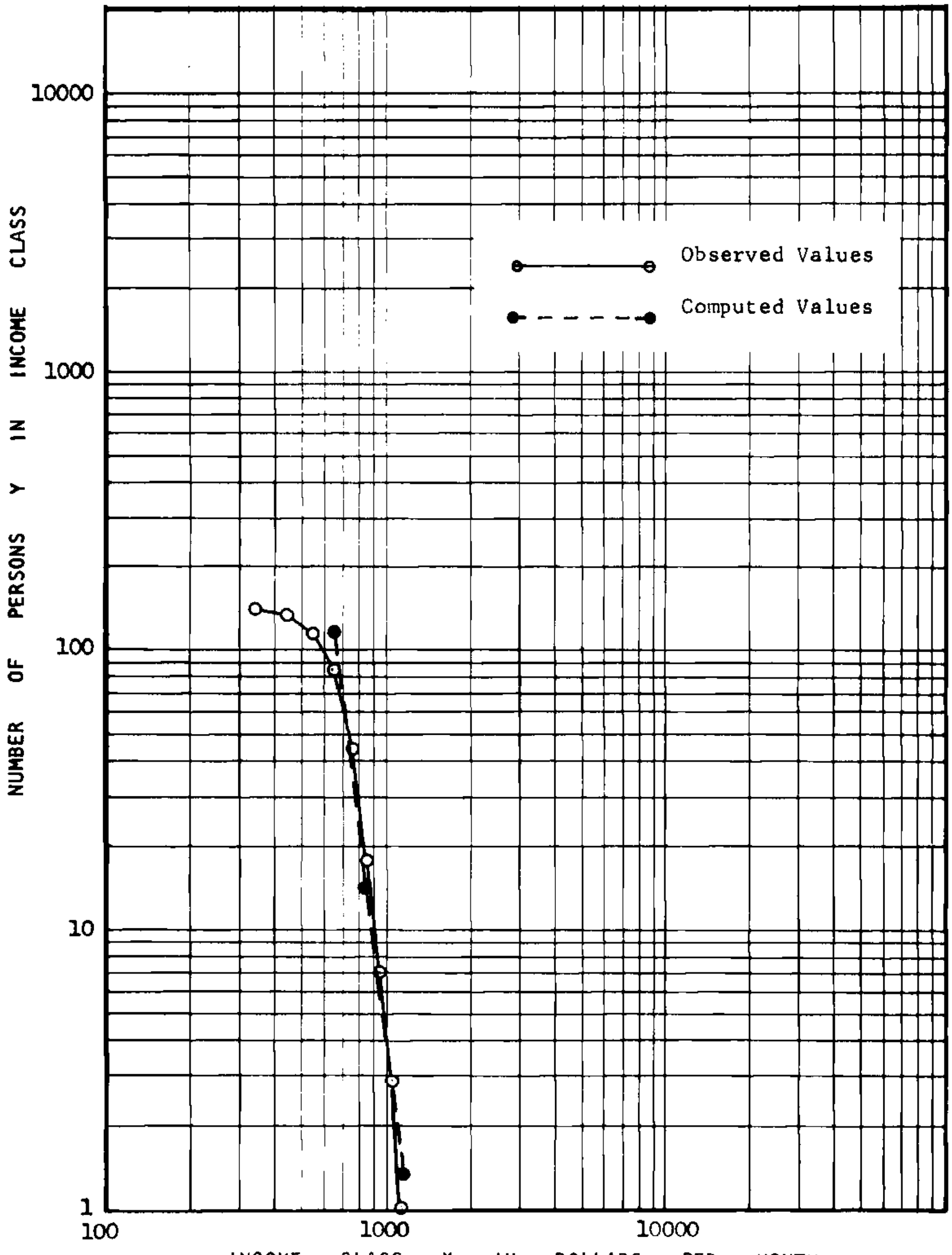

Figure 29. Pareto Curve Fitted to Data of Company F in Range R_u

Table 10. Cumulative Frequency Tabulation for Data of Company F (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 850	17	2.92942	1.23045	3.60451	8.58150	22.79804	1.15121	14.16
950	7	2.97772	0.84510	2.51647	8.86682	23.17393	0.77532	5.96
1,050	3	3.02119	0.47712	1.44147	9.12759	23.51223	0.43702	2.74
1,150	1	3.06070			9.36788	23.81971	0.12954	1.35
Totals		17.67700	6.12531	17.71285	52.12222			

Table 11. Cumulative Frequency Tabulation for Data of Industry I

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	8,137							
350	8,134							
450	7,781							

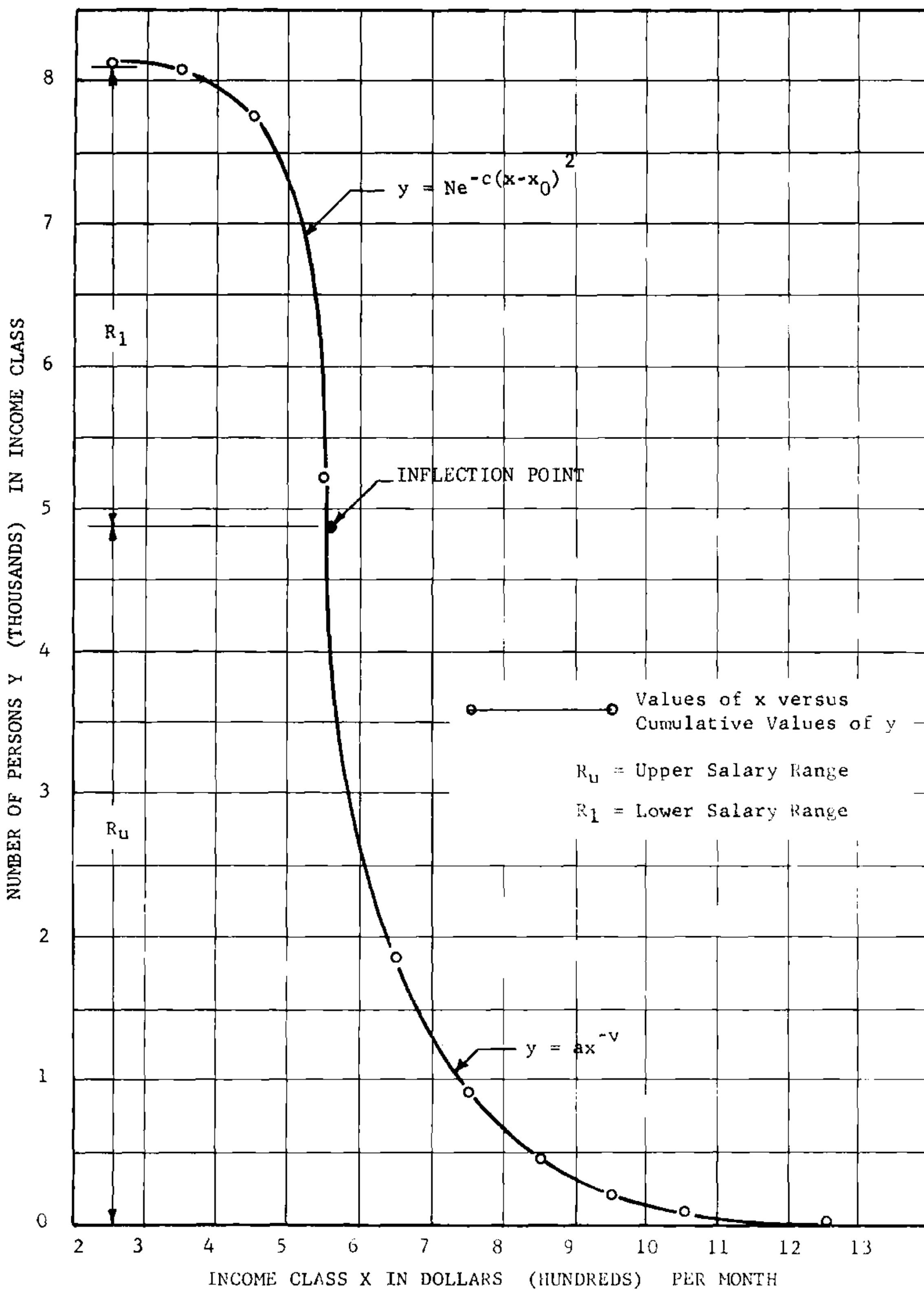

Figure 30. Distribution of Incomes for Data of Industry I

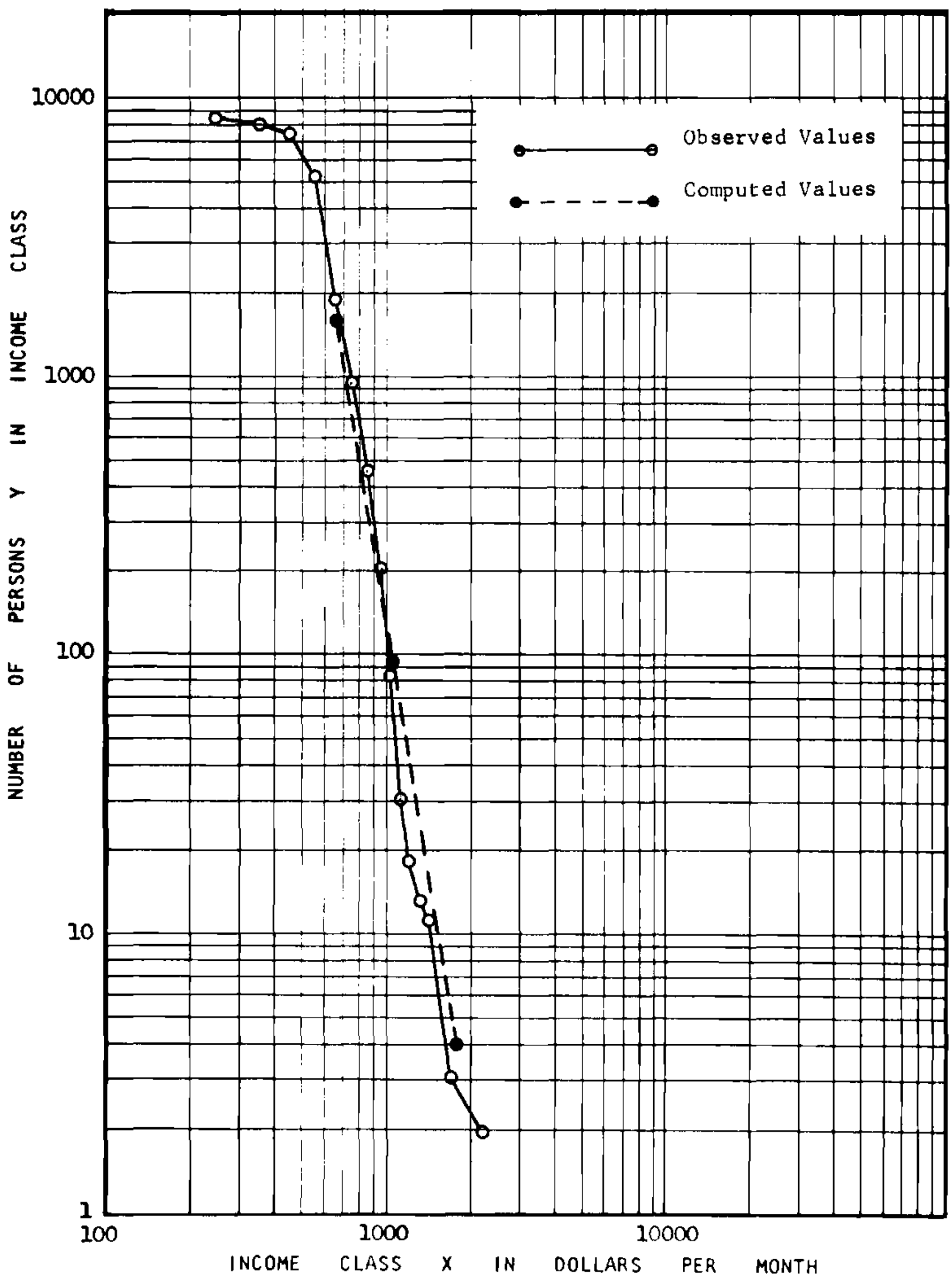

Figure 31. Pareto Curve Fitted to Data of Industry I in Range R_u

Table 11. Cumulative Frequency Tabulation for Data of Industry I (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 550	5,212							
650	1,869	2.81291	3.27161	9.20275	7.91246	17.06187	3.21382	1636.10
750	925	2.87506	2.96614	8.52783	8.26597	17.43884	2.83685	686.84
850	458	2.92942	2.66087	7.79481	8.58150	17.76857	2.50712	321.46
950	198	2.97772	2.29667	6.83884	8.86682	18.06154	2.21415	163.74
1,050	84	3.02119	1.92428	5.81362	9.12759	18.32521	1.95048	89.22
1,150	30	3.06070	1.47712	4.52102	9.36788	18.56486	1.71083	51.38
1,250	18	3.09691	1.25527	3.88746	9.59085	18.78449	1.49120	30.98
1,350	13	3.13033	1.11394	3.48700	9.79897	18.9872	1.28849	19.43
1,450	11	3.16137	1.04139	3.29222	9.99426	19.17548	1.10021	12.59
1,550	10	3.19033	1.00000	3.19033	10.17821	19.35114	0.92455	8.41
1,750	3	3.24304	0.47712	1.54732	10.51731	19.67085	0.60484	4.02

Table 11. Cumulative Frequency Tabulation for Data of Industry I (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$2,250	2	3.35218	0.30103	1.00911	11.23711	20.33285		
Totals		36.85116	19.78544	59.11231	113.43893			

Table 12. Comparison of Computed Values Versus Actual Values For Data of Company J Based on Method III

Income Class Mark (x)	$y = Ne^{-c(x-x_0)^2}$	Actual Values of y in Range R_1
$450	1128.00	1128
550	1095.29	1126
650	990.38	990
750	835.85	828
850	564.39	651

Table 13. Comparison of Computed Values Versus Actual Values For Data of Company E Based on Method III

Income Class Mark (x)	$y = Ne^{-c(x-x_0)^2}$	Actual Values of y in Range R_1
$250	1397.00	1397
350	1342.52	1394
450	1179.07	1065

NOTE: After determining c from the inflection point, it was divided by 2 because the resulting constant improved the fitted curve.

Table 14. Comparison of Computed Versus Actual Values For Data of Company F Based on Method III

Income Class Mark (x)	$y = Ne^{-c(x-x_0)^2}$	Actual Values of y in Range R_1
\$350	139.00	139
450	136.22	137
550	125.66	118
650	111.48	89

Table 15. Comparison of Computed Versus Actual Values For Data of Industry I Based on Method III

Income Class Mark (x)	$y = Ne^{-c(x-x_0)^2}$	Actual Values of y in Range R_1
\$250	8137.00	8137
350	7974.26	8134
450	7289.13	7781
550	6337.09	5212

the tabulated and graphical results pertaining to data of the above companies. A summary for computed values of y using Method III versus observed values is presented in Tables 12 through 15.

Remarks

Analytical investigations of income using the three methods indicate that Method III gives the best results. The author feels that the empirical formula suggested in Method III is an adequate and satisfactory description of income data in range R_1 for all practical purposes. However, it is not claimed that this is the best representation of income data in mathematical terms. An exhaustive study of income data at a much higher mathematical level is beyond the scope of this study.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

It must be kept in mind that Pareto distribution is but an "empirical" and not a rational law of the distribution of incomes. The assumption made is that some disturbing forces cause a distortion of the fundamental income curve, which is supposed to be the result of a large number of unbiased forces (25). The disturbing forces are a composite of elements of skill, technological advances, government legislation, union pressure, market demand and supply considerations, and personal circumstances favoring the acquisition of income.

In applying the Pareto distribution to income data of the participating companies we have been concerned with wage and salary income in various manufacturing industries. After a detailed examination of income statistics for the 15 establishments out of 211 contacted, the following conclusions have been reached:

(1) From the practical standpoint, it can be stated that the Pareto formula generally fits well limited portions of the observed income curve, with different values of v. However, Pareto's law failed to represent with sufficient accuracy the actual distribution of the income data over the whole range.

(2) The Pareto law is not extendable to incomes in the lower wage and salary range R_1. Extension of logarithmic straight-line to range R_1 results in the absurd conclusion of an infinite number of

persons having incomes just above the $250 amount.

(3) The value of slope v can be used as an index of inequality of incomes in comparing a particular company with other companies and the industry as a whole. The greater the value of v, the less the inequality of the distribution of income for a given company; and, the smaller the value of v, the greater the spread of distribution for incomes.

(4) The Pareto law can be usefully applied as a quantitative method of comparison in analyzing the structure of salaries in the manufacturing industries.

(5) The significance of Pareto's law lies in the fact that it suggests the existence of an "order" in the distribution of wage and salary incomes. The theory of statistical regularities in human actions can be extended to the distribution of incomes in industrial organizations.

(6) As a useful tool econometric methods of analysis are invaluable to the industrial engineer, since they permit quantitative study of the structure of economic and industrial systems.

The following recommendations are made as a result of this study:

(1) Statistical income data should be collected through carefully prepared questionnaires from a larger fraction of the companies making up a particular industry, and a similar analysis should be undertaken with these data.

(2) It may be argued that if the Pareto distribution cannot be applied to income statistics of the 15 companies studied, it is not necessarily due to the inherent defects of the "law" but to the in-

accuracies of the statistics examined. Therefore, future investigations must be aimed at collecting representative data which is accurate and complete.

(3) It is widely accepted that the fundamental causes of inequality in income distribution are the influence of inherited ability and environmental factors. It would be of interest to ascertain the extent these influences affect the form of the income curve for various income classes.

(4) It is recommended that more studies be undertaken to ascertain the existence of income distribution "patterns" by use of empirical formulas other than the Pareto distribution.

A P P E N D I X

PART I

Excerpts from Letters

We have completed your Salary and Wage Questionnaire to reflect the distribution of our employees by payroll category and by salary received.

We are especially interested in your results and would appreciate receiving a copy of your study.

At your request of November 7, we have completed the enclosed forms reflecting distribution of base monthly salaries by payroll category for the . . .

Any additional information required, or classification of the submitted material desired may be obtained by contacting the undersigned.

I am enclosing the completed questionnaire forms which you recently addressed to me. I trust that the information as supplied will be helpful to you in the preparation of your project.

We wish you success in the completion of your studies and would be most interested to learn of the conclusions you may reach from this investigation.

Enclosed is the data you requested, covering a typical location . . .

We would appreciate receiving a copy of the report which will be prepared.

Please find enclosed survey which you forwarded to us for completion.

I would like to point out that the different classifications could well be misleading in that it is strictly our interpretation of what you wanted in the survey because in actuality we have nothing similar to this breakdown in our salary structure.

I regret the necessity of advising you that the release of such information is not in consonance with company policy.

We are sorry that we are unable to comply with your request for salary information. As a practical measure we have made it our policy to provide salary information only on a reciprocal basis with other industrial or institutional organizations.

The information which you requested in reference to salaries is termed as "Classified Material" with the . . . , and we would not be able to give this data at this time.

I must start this letter by indicating that we regretfully cannot furnish the information you desire.

We are very interested in the subject matter of your research project. We ultimately hope to utilize this technique in our comparison efforts as we would be most interested in the comparative profiles of our company and the industry as a whole.

We find that unfortunately, our records are not set up in such a way that we can readily supply the data you request. It would require a monumental effort at this year-end closing time to pull out the information and report it satisfactorily to you.

PART II

Sample Calculations for Chapter IV

Company A:

$$N \log a - v \sum \log x = \sum \log y$$

$$(\sum \log x) \log a - v \sum (\log x)^2 = \sum (\log x)(\log y)$$

$$7 \log a - 20.41736\ v = 14.23069$$

$$20.41736 \log a - 59.63179\ v = 40.82328$$

$$142.92152 \log a - 416.86859\ v = 290.55312$$

$$\underline{142.92152 \log a - 417.42253\ v} = \underline{285.76296}$$

$$0 \quad + \quad 0.55394\ v = \quad 4.79016$$

$$v = \frac{4.79016}{0.55394} = 8.64743$$

$$7 \log a = 14.23069 + (20.41736)(8.64743)$$

$$= 14.23069 + 176.55769$$

$$\log a = \frac{190.78838}{7} = 27.44517$$

The computed equation for Company A comes out to be:

$$\log y = 27.44517 - 8.64743 \log x .$$

<u>Company O:</u>

$$22 \log a - 68.4622 \ v = 36.77517$$

$$64.46222 \log a - 216.01190 \ v = 106.95554$$

$$1506.16884 \log a - 4687.07557 \ v = 2517.70978$$

$$\underline{1506.16884 \log a - 4752.26180 \ v} = \underline{2353.02188}$$

$$0 \quad + \quad 65.18623 \ v = 164.68790$$

$$v = \frac{164.68790}{65.18623} = 2.52642$$

$$22 \log a = 36.77517 + (68.46222)(2.52642)$$

$$= 36.77517 + 172.96432$$

$$\log a = \frac{209.73949}{22} = 9.53361$$

The computed equation for Company O comes out to be:

$$\log y = 9.53361 - 2.52642 \log x.$$

<u>INDUSTRY I:</u>

$$16 \log a - 47.18674 \ v = 35.21424$$

$$47.18674 \log a - 140.21043 \ v = 98.94711$$

$$\begin{array}{rcl} 754.98784 \log a - 2226.58843\ v & = & 1661.64519 \\ \underline{754.98784 \log a - 2243.36688\ v} & = & \underline{1583.15376} \\ 0 \qquad + \quad 16.77845\ v & = & 78.49143 \end{array}$$

$$v = \frac{78.49143}{16.77845} = 4.67811$$

$$16 \log a = 35.21424 + (47.18674)(4.67811)$$

$$= 35.21424 + 220.74476$$

$$\log a = \frac{255.95900}{16} = 15.99744$$

The computed equation for Company I comes out to be:

$$\log y = 15.99744 - 4.67811 \log x.$$

PART III

Calculations for the Method of Finite Differences

It can be shown that the relationship

$$\Delta_1 \log y = b \log 3 + 2\ cx \log e$$

is true. Using data of Table 6 we proceed as follows:

$$-3.209 = 3\ b \log 3 + 2c\ (250+350+450) \log e, \qquad (1)$$

$$-5.617 = 3\ b \log 3 + 2c\ (550+650+750) \log e. \qquad (2)$$

Simplifying (1) and (2) we get

$$-3.209 = 1.431\ b + 913.5\ c \qquad (3)$$

$$-5.617 = 1.431\ b + 1696.5\ c. \qquad (4)$$

Solving (3) and (4) simultaneously,

$$c = -\frac{2.408}{783} = -0.00308 \approx -0.0031.$$

Substituting the value of c in (3) and solving for b, we get b = -0.264.

From the relation (where, $\sum \log y = 17.031$; $N = 18$; $\sum \log x =$ 20,950;)

$$\sum \log y = N \log a - b \left(\sum \log x \right) - c \left(\sum x \right) \log e$$

the value of the parameter a is found to be 1.13. Then the general equation (1) in Chapter V becomes

$$y = 1.13\ x^{-0.264}\ e^{-0.0031x}. \tag{5}$$

Substituting values of x in (5) above gives very small values of y which is not consistent with the observed values.

Calculations for the Method of Selected Points

We are to determine the parameters a, b, and c in the general equation $y = ax^b e^{cx}$ using the method of selected points. It can be shown that a, b, and c are expressed by the expressions below (23):

$$b = \frac{\ln(y_1/y_2) - [(x_1-x_2)/(x_1-x_3)]\ \ln(y_1/y_3)}{\ln(x_1/x_2) - [(x_1-x_2)/(x_1-x_3)]\ \ln(x_1/x_3)},$$

$$c = \frac{\ln(y_1/y_2) - b\ \ln\ (x_1/x_2)}{(x_1 - x_2)}, \text{ and}$$

$$a = y_1 x_1^{-b} e^{-cx_1}.$$

The widely spaced selected points were $x_1 = 35$, $x_2 = 55$, $x_3 = 95$, $y_1 = 1394$, $y_2 = 328$, and $y_3 = 31$.

From the above relations a, b, and c were found to be 5.17, -0.03549, and -1.64, respectively. The general equation (1) of Chapter

V takes the form

$$y = 5.17\ x^{-1.64} e^{-0.036x} \quad . \tag{6}$$

Again substitution of values for x results in corresponding values of y which differ significantly from the observed values.

PART IV

Table 16. Cumulative Frequency Tabulation for Data of Company A

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	2,946							
450	2,933							
550	2,186	2.74036	3.33965	9.15184	7.50957	23.69708	3.74809	5598.60
650	564	2.81291	2.75128	7.73910	7.91246	24.32444	3.12073	1320.50
750	330	2.87506	2.51851	7.24087	8.26597	24.86188	2.58329	383.08
850	189	2.92942	2.27646	6.66871	8.58150	25.33195	2.11322	129.78
950	79	2.97772	1.89763	5.65061	8.86682	25.74963	1.69554	49.61
1,050	28	3.02119	1.44716	4.37215	9.12759	26.12553	1.31964	20.88
1,150	1	3.06070			9.36788	26.46719	0.97798	9.51
Totals		20.41736	14.23069	40.82328	59.63179			

Table 17. Cumulative Frequency Tabulation for Data of Company B

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	2,305							
450	2,298							
550	1,315	2.74036	3.11893	8.54699	7.50957	19.87301	3.32514	2114.20
650	564	2.81291	2.75128	7.73910	7.91246	20.39914	2.79901	629.52
750	292	2.87506	2.46538	7.08812	8.26597	20.84985	2.34830	223.00
850	145	2.92942	2.16137	6.33156	8.58150	21.24407	1.95408	89.97
950	61	2.97772	1.78533	5.31621	8.86682	21.59434	1.60381	40.16
1,050	25	3.02119	1.39794	4.22344	9.12759	21.90958	1.28857	19.43
1,150	8	3.06070	0.90309	2.76409	9.36788	22.19611	1.00204	10.05
1,250	3	3.09691	0.47712	1.47760	9.59085	22.45870	0.73945	5.49
Totals		23.51427	15.06044	43.48711	69.22264			

Table 18. Cumulative Frequency Tabulation for Data of Company C

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 450	296	2.65321	2.47129	6.55685	7.03952	14.87883	2.66150	458.67
550	258	2.74036	2.41162	6.06871	7.50957	15.36756	2.17277	148.86
650	59	2.81291	1.77085	4.98124	7.91246	15.77441	1.76592	58.34
750	26	2.87506	1.41497	4.06812	8.26597	16.12293	1.41740	26.15
850	16	2.92942	1.20412	3.52737	8.58150	16.42778	1.11255	12.96
950	5	2.97772	0.69897	2.08134	8.86682	16.69864	0.84169	6.95
						16.94241	0.59792	3.96
						17.16398	0.37635	2.38
						17.36704	0.17329	1.49
Totals		16.98868	9.97182	27.82363	48.17584			

Table 19. Cumulative Frequency Tabulation for Data of Company D

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	1,054	2.54407	3.02284	7.69032	6.47229	12.09123	3.45898	2877.30
450	1,052	2.65321	3.02202	8.01805	7.03952	12.60994	2.94027	871.02
550	1,007	2.74036	3.00303	8.22938	7.50957	13.02414	2.52607	335.79
650	439	2.81291	2.64246	7.43300	7.91246	13.36895	2.18126	151.80
750	134	2.87506	2.12710	6.11554	8.26597	13.66433	1.88588	76.89
850	35	2.92942	1.54407	4.52323	8.58150	13.92268	1.62753	42.42
950	15	2.97772	1.17609	3.50207	8.86682	14.15224	1.39797	25.00
1,050	9	3.02119	0.95424	2.88294	9.12759	14.35884	1.19137	15.54
1,150	6	3.06070	0.77815	2.38168	9.36789	14.54662	1.00359	10.08
1,250	3	3.09691	0.47712	1.47760	9.59085	14.71872	0.83149	6.78
1,550	2	3.19033	0.30103	0.96039	10.17821	15.16271	0.38750	2.44

Table 19. Cumulative Frequency Tabulation for Data of Company D (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$2,250	1	3.35218			11.23711	15.93194		
Totals		35.25406	19.05015	53.21420	104.14978			

Table 20. Cumulative Frequency Tabulation for Data of Company E

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	1,397	2.39794	3.14520	7.54200	5.75012	8.66793	3.61555	4126.20
350	1,394	2.54407	3.14426	7.99922	6.47229	9.19615	3.08733	1222.70
450	1,065	2.65321	3.02735	8.03220	7.03952	9.59066	2.69282	492.97
550	328	2.74036	2.51587	6.89439	7.50957	3.90569	2.37779	238.67
650	154	2.81291	2.18752	6.15330	7.91246	10.16794	2.11554	130.48
750	101	2.87506	2.00432	5.76254	8.26597	10.39259	1.89089	77.79

Table 20. Cumulative Frequency Tabulation for Data of Company E (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 850	56	2.92942	1.74819	5.12118	8.58150	10.58909	1.69439	49.48
950	31	2.97772	1.49136	4.44085	8.86682	10.76368	1.51980	33.10
1,050	19	3.02119	1.27875	3.86335	9.12759	10.92082	1.36266	23.05
1,150	14	3.06070	1.14613	3.50796	9.36788	11.06364	1.21984	16.59
1,250	12	3.09691	1.07918	3.34212	9.59085	11.19452	1.08896	12.27
1,350	11	3.13033	1.04139	3.25989	9.79897	11.31533	0.96815	9.29
1,450	9	3.16137	0.95424	3.01671	9.99426	11.42753	0.85594	7.18
1,550	8	3.19033	0.90309	2.88116	10.17821	11.53221	0.75127	5.64
1,750	2	3.24304	0.30103	0.97625	10.51731	11.72275	0.56073	3.64
2,250	1	3.35218			11.23711	12.11726	0.16622	1.47
Totals		47.18674	25.96788	72.79312	140.21043			

Table 21. Cumulative Frequency Tabulation for Data of Company F

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	139							
450	137							
550	118	2.74036	2.07188	5.67770	7.50957	17.96331	2.32737	212.51
650	89	2.81291	1.94939	5.48346	7.91246	18.43888	1.85180	71.09
750	42	2.87506	1.62325	4.66694	8.26597	18.84628	1.44440	27.82
850	17	2.92942	1.23045	3.60451	8.58150	19.20261	1.08807	12.25
950	7	2.97772	0.84510	2.51647	8.86682	19.51922	0.77146	5.91
1,050	3	3.02119	0.47712	1.44147	9.12759	19.80417	0.48651	3.07
1,150	1	3.06070			9.36788	20.06316	0.22752	1.69
Totals		20.41736	8.19719	23.39055	59.63179			

Table 22. Cumulative Frequency Tabulation for Data of Company G

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	4,440							
350	4,433							
450	4,382	2.65321	3.64167	9.66212	7.03952	9.83495	3.79377	6219.70
550	3,794	2.74036	3.57910	9.80802	7.50957	10.15799	3.47073	2956.20
650	1,257	2.81291	3.09934	8.71817	7.91246	10.42692	3.20180	1591.50
750	988	2.87506	2.99476	8.61012	8.26597	10,65730	2.97142	936.32
850	655	2.92942	2.81624	8.24995	8.58150	10.85880	2.76992	588.74
950	449	2.97772	2.65225	7.89766	8.86682	11.03784	2.59088	389.83
1,050	327	3.02119	2.51455	7.59693	9.12759	11.19898	2.42974	268.99
1,150	242	3.06070	2.38382	7.29616	9.36789	11.34543	2.28329	192.00
1,250	167	3.09691	2.22272	6.88356	9.59085	11.47966	2.14906	140.95
1,350	118	3.13033	2.07188	6.48567	9.79897	11.60354	2.02518	105.97
1,450	72	3.16137	1.85733	5.87171	9.99426	11.71860	1.91012	81.31
1,550	50	3.19033	1.69897	5.42028	10.17821	11.82595	1.80277	63.50

Table 22. Cumulative Frequency Tabulation for Data of Company G (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 1,650	37	3.21748	1.56820	5.04565	10.35218	11.92659	1.70213	50.37
21,250	35							
Totals		38.86699	33.10083	97.54600	116.58579			

Table 23. Cumulative Frequency Tabulation for Data of Company H

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	4,301							
450	4,289							
550	3,715							
650	3,377	2.81291	3.52853	9.92544	7.91246	19.14194	3.61773	4147.00
750	724	2.87506	2.85974	8.22192	8.26597	19.56487	3.19480	1566.10

Table 23. Cumulative Frequency Tabulation for Data of Company H (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 850	580	2.92942	2.76343	8.09525	8.58150	19.93479	2.82488	668.16
950	362	2.97772	2.55871	7.61912	8.68682	20.26347	2.49620	313.48
1,050	226	3.02119	2.35411	7.11221	9.12759	20.55929	2.20038	158.63
1,150	150	3.06070	2.17609	6.66036	9.36788	20.82816	1.93151	85.41
1,250	86	3.09691	1.93450	5.99097	9.59085	21.07457	1.68510	48.43
1,350	46	3.13033	1.66276	5.20499	9.79897	21.30199	1.45768	28.69
1,450	33	3.16137	1.51851	4.80057	9.99426	21.51322	1.24645	17.64
1,550	16	3.19033	1.20412	3.84154	10.17821	21.71029	1.04938	11.20
1,650	5	3.21748	0.69897	2.24892	10.35218	21.89505	0.86462	7.32
1,750	1	3.24304			10.51731	22.06898	0.69069	4.91
Totals		36.71646	23.25947	69.72129	112.55400			

Table 24. Cumulative Frequency Tabulation for Data of Company I

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	5,372	2.54407	3.73014	9.48974	6.47229	6.75245	3.48814	3077.10
450	4,352	2.65321	3.63869	9.65421	7.03952	7.04212	3.19847	1579.30
550	1,888	2.74036	3.27600	8.97742	7.50957	7.27344	2.96715	927.15
650	821	2.81291	2.91434	8.19778	7.91246	7.46560	2.77499	595.65
750	608	2.87506	2.78390	8.00388	8.26597	7.63096	2.60963	407.04
850	408	2.92942	2.61066	7.64772	8.58150	7.77524	2.46535	291.98
950	271	2.97772	2.43297	7.24470	8.86682	7.90344	2.33715	217.35
1,050	122	3.02119	2.08636	6.30329	9.12759	8.01881	2.22178	166.64
1,150	81	3.06070	1.90849	5.84132	9.36788	8.12368	2.11691	130.89
1,250	64	3.09691	1.80618	5.59358	9.59085	8.21979	2.02080	104.91
1,350	40	3.13033	1.60206	5.01498	9.79897	8.30849	1.93210	85.53
1,550	32	3.19033	1.50515	4.80193	10.17821	8.46774	1.77285	59.27
1,650	31	3.21748	1.49136	4.79842	10.35218	8.53980	1.70079	50.21
1,750	25	3.24304	1.39794	4.53363	10.51731	8.60775	1.63284	42.94

Table 24. Cumulative Frequency Tabulation for Data of Company I (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 1,850	19	3.26717	1.27875	4.17789	10.67440	8.67169	1.56890	37.06
1,950	16	3.29003	1.20412	3.96159	10.82430	8.73237	1.50822	32.23
2,250	14	3.35218	1.14613	3.84203	11.23711	8.89732	1.34327	22.04
2,750	9	3.43933	0.95424	3.28195	11.82693	9.12864	1.11195	12.94
3,250	7	3.51188	0.84510	2.96789	12.33330	9.32120	0.91939	8.31
3,750	6	3.57403	0.77815	2.78113	12.77369	9.48616	0.75443	5.68
5,250	4	3.72016	0.60206	2.23976	13.83959	9.87401	0.36658	2.33
7,250	2	3.86034	0.30103	1.16208	14.90222			
11,250	1	4.05115			16.41182			
Totals		73.55904	40.29382	120.51692	238.40448			

Table 25. Cumulative Frequency Tabulation for Data of Company J

Income Class Mark (x)	Cumulative Number of Persons	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log x	Computed y Cumulative
$ 450	1,128							
550	1,126							
650	990							
750	828	2.87506	2.91803	8.38951	8.26597	19.41045	3.37646	2379.40
850	651	2.92942	2.81358	8.24216	8.58150	19.77745	3.00946	1022.00
950	508	2.97772	2.70586	8.05729	8.86682	20.10354	2.68337	482.36
1,050	377	3.02119	2.57634	7.78361	9.12759	20.39702	2.38989	245.41
1,150	250	3.06070	2.39794	7.33938	9.36788	20.66377	2.12314	132.79
1,250	162	3.09691	2.20952	6.84269	9.59085	20.90825	1.87868	75.63
1,350	97	3.13033	1.98677	6.21925	9.79897	21.13386	1.65305	44.98
1,450	56	3.16137	1.74819	5.52668	9.99426	21.34342	1.44349	27.77
1,550	26	3.19033	1.41497	4.51422	10.17821	21.53894	1.24797	17.70
1,650	7	3.21748	0.84510	2.71909	10.35218	21.72224	1.06467	11.61
1,750	2	3.24304	0.30103	1.97626	10.51731	21.89480	0.89211	7.80

Table 25. Cumulative Frequency Tabulation for Data of Company J (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$2,250	1	3.35218			11.23711	22.63164	0.15527	1.43
Totals		37.25577	21.91733	66.61014	115.87865			

Table 26. Cumulative Frequency Tabulation for Data of Company K

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	1,641							
350	1,593							
450	1,535	2.65321	3.18611	8.45342	7.03952	19.87437	3.21653	1646.40
550	332	2.74036	2.52114	6.90883	7.50957	20.52719	2.56371	366.20
650	97	2.81291	1.98677	5.58861	7.91246	21.07064	2.02026	104.78
750	45	2.87506	1.65321	4.75308	8.26597	21.53618	1.55472	35.87

Table 26. Cumulative Frequency Tabulation for Data of Company K (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log x	Computed y Cumulative
$ 850	21	2.92942	1.32222	3.87334	8.58150	21.94338	1.14752	14.05
950	6	2.97772	0.77815	2.31711	8.86682	22.30518	0.78572	6.11
1,050	2	3.02119	0.30103	0.90947	9.12759	22.63080	0.46010	2.88
Totals		20.00987	11.74863	32.80386	57.30343			

Table 27. Cumulative Frequency Tabulation for Data of Company L

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log x	Computed y Cumulative
$ 250	5,856							
350	5,803							
450	5,621							
550	2,789	2.74036	3.44545	9.44177	7.50957	14.95069	3.78274	6063.80

Table 27. Cumulative Frequency Tabulation for Data of Company L (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 650	1,621	2.81291	3.20978	9.02882	7.91246	15.34651	3.38692	2437.40
750	1,134	2.87506	3.05461	8.78219	8.26597	15.68558	3.04785	1116.50
850	809	2.92942	2.90795	8.51861	8.58150	15.98215	2.75128	564.00
950	362	2.97772	2.55871	7.61912	8.86682	16.24567	2.48776	307.44
1,050	258	3.02119	2.41162	7.28596	9.12759	16.48283	2.25060	178.08
1,150	155	3.06070	2.19033	6.70394	9.36788	16.69838	2.03505	108.41
1,250	109	3.09691	2.03743	6.30974	9.59085	16.89594	1.83749	68.79
1,350	54	3.13033	1.73239	5.42295	9.79897	17.07827	1.65516	45.20
1,450	36	3.16137	1.55630	4.92004	9.99426	17.24761	1.48582	30.61
1,550	30	3.19033	1.47712	4.71250	10.17821	17.40561	1.32782	21.27
1,650	19	3.21748	1.27875	4.11435	10.35218	17.55373	1.17970	15.13
1,850	8	3.26717	0.90309	2.95055	10.67440	17.82483	0.90860	8.10
1,950	4	3.29003	0.60206	1.98080	10.82430	17.94955	0.78388	6.08

Table 27. Cumulative Frequency Tabulation for Data of Company L
(Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$2,250	1	3.35218			11.23711	18.28862	0.44481	2.78
Totals		46.12316	29.36559	87,79134	142.28207			

Table 28. Cumulative Frequency Tabulation for Data of Company M

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 350	10,343							
450	9,980							
550	8,296	2.74036	3.91887	10.73912	7.50957	13.50617	4.16194	14519.00
650	4,955	2.81291	3.69504	10.39382	7.91246	13.86374	3.80437	6373.40
750	3,151	2.87506	3.49845	10.05825	8.26597	14.17005	3.49806	3148.20
850	2,226	2.92942	3.34753	9.80632	8.58150	14.43797	3.23014	1698.80

Table 28. Cumulative Frequency Tabulation for Data of Company M (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 950	1,465	2.97772	3.16584	9.42699	8.86682	14.67602	2.99209	981.94
1,050	920	3.02119	2.96379	8.95417	9.12759	14.89027	2.77784	599.57
1,150	559	3.06070	2.74741	8.40900	9.36788	15.08500	2.58311	382.92
1,250	356	3.09691	2.55145	7.90161	9.59085	15.26346	2.40465	253.89
1,350	212	3.13033	2.32634	7.28221	9.79897	15.42818	2.23993	173.75
1,450	119	3.16137	2.07555	6.56158	9.99426	15.58116	2.08695	122.17
1,550	73	3.19033	1.86332	5.94461	10.17821	15.72389	1.94422	87.95
1,650	24	3.21748	1.38021	4.44081	10.35218	15.85770	1.81041	64.63
Totals		36.21378	33.53380	99.91849	109.54626			

Table 29. Cumulative Frequency Tabulation for Data of Company N

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	1,940	2.39794	3.28780	7.88395	5.75012	8.20786	3.53351	3416.00
350	1,206	2.54407	3.08135	7.83917	6.47229	8.70805	3.0332	1079.70
450	890	2.65321	2.94939	7.82535	7.03952	9.08162	2.65975	456.83
550	484	2.74036	2.68485	7.35746	7.50957	9.37992	2.36145	229.85
650	226	2.81291	2.35411	6.62190	7.91246	9.62825	2.11312	129.75
750	104	2.87506	2.01703	5.79908	8.26597	9.84099	1.90038	79.50
850	47	2.92942	1.67210	4.89828	8.58150	10.02705	1.71432	51.80
950	18	2.97772	1.25527	3.73784	8.86682	10.19238	1.54899	35.40
1,050	11	3.02119	1.04139	3.14624	9.12759	10.34117	1.40020	25.13
1,250	9	3.09691	0.95424	2.95520	9.59085	10.60035	1.14102	13.83
1,350	6	3.13033	0.77815	2.43587	9.79897	10.71474	1.02663	10.63
1,850	4	3.26717	0.60206	1.96703	10.67440	11.18313	0.55824	3.62
2,250	2	3.35218	0.30103	1.00911	11.23711	12.09023		

Table 29. Cumulative Frequency Tabulation for Data of Company N (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$3,250	1	3.51188			12.33330			
Totals		41.31035	22.97877	63.47648	123.16047			

Table 30. Cumulative Frequency Tabulation for Data of Company O

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	9,670	2.39794	3.98543	9.55682	5.75012	6.05820	3.47541	2988.20
350	3,722	2.54407	3.57078	9.08431	6.47229	6.42739	3.10622	1277.10
450	786	2.65321	2.89542	7.68216	7.03952	6.70312	2.83049	676.85
550	437	2.74036	2.64048	7.23587	7.50957	6.92330	2.61031	407.67
650	208	2.81291	2.31806	6.52049	7.91246	7.10659	2.42702	267.31
750	136	2.87506	2.13354	6.13406	8.26597	7.26361	2.27000	186.21

Table 30. Cumulative Frequency Tabulation for Data of Company O

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 850	97	2.92942	1.98677	5.82008	8.58150	7.40095	2.13266	135.73
950	69	2.97772	1.83885	5.47558	8.86682	4.52297	2.01064	102.48
1,050	53	3.02119	1.72428	5.20938	9.12759	7.63279	1.90082	79.58
1,150	45	3.06070	1.65321	5.05998	9.36788	7.73261	1,80100	63.24
1,250	39	3.09691	1.59106	4.92737	9.59085	7.82410	1.70951	51.228
1,350	30	3.13033	1.47712	4.62387	9.79897	7.90853	1.62508	42.177
1,450	24	3.16137	1.38021	4.36335	9.99426	7.98695	1.54666	35.210
1,550	22	3.19033	1.34242	4.28276	10.17821	8.06011	1.47350	29.751
1,650	16	3.21748	1.20412	3.87423	10.35218	8.12871	1.40490	25.404
1,850	15	3.26717	1.17609	3.84249	10.67440	8.25424	1.27937	19.027
1,950	10	3.29003	1.00000	3.29003	10.82430	8.31200	1.22161	16.658
2,250	9	3.35218	0.95424	3.19878	11.23711	8.46901	1.06460	11.604
2,750	8	3.43933	0.90309	3.10602	11.82899	8.68919	0.84442	6.9891
4,250	5	3,62839	0.69897	2.53614	13.16521	9.16684	0.36677	2.3268

Table 30. Cumulative Frequency Tabulation for Data of Company O (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$5,750	2	3.75967	0.30103	1.13177	14.13512	9.49851	0.03510	1.0842
8,250	1	3.91645			15.33858			
Totals		68.46222	36.77517	106.95554	216.01190			

Table 31. Cumulative Frequency Tabulation for Data of Industry I

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	8,137	2.39794	3.91046	9.37705	5.75012	11.21783	4.77961	60202.00
350	8,134	2.54407	3.91030	9.94808	6.47229	11.90144	4.09600	12474.00
450	7,781	2.65321	3.89104	10.32375	7.03952	12.41201	3.58543	3849.80
550	5,212	2.74036	3.71700	10.18592	7.50957	12.81971	3.17773	1505.70
650	1,869	2.81291	3.27161	9.20275	7.91246	13.15910	2.83834	689.19

Table 31. Cumulative Frequency Tabulation for Data of Industry I (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 750	925	2.87506	2.96614	8.52783	8.26597	13.44985	2.54759	352.85
850	458	2.92942	2.66087	7.79481	8.58150	13.70415	2.29329	196.47
950	198	2.97772	2.9667	6.83884	8.86682	13.93010	2.06734	116.77
1,050	84	3.02119	1.92428	5.81362	9.12759	14.13346	1.86398	73.11
1,150	30	3.06070	1.47712	4.52102	9.36788	14.31829	1.67915	47.77
1,250	18	3.09691	1.25527	3.88746	9.59085	14.48769	1.50975	32.34
1,350	13	3.13033	1.11394	3.48700	9.79897	14.64403	1.35341	22.56
1,450	11	3.16137	1.04139	3.29222	9.99426	14.78924	1.20820	16.15
1,550	10	3.19033	1.00000	3.19033	10.17821	14.92472	1.07272	11.82
1,750	3	3.24304	0.47712	1.54732	10.51731	15.17130	0.82614	6.69
2,250	2	3.35218	0.30103	1.00911	11.23711	15.68187	0.31557	2.07
Totals		47.18674	35.21424	98.94711	140.21043			

Table 32. Cumulative Frequency Tabulation for Data of Industry II

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	8,706							
350	8,699							
450	8,636	2.65321	3.93631	10.44386	7.03952	15.20151	4.50079	31680.00
550	7,474	2.74036	3.87355	10.61492	7.50957	15.70084	4.00146	10034.00
650	4,599	2.81291	3.66266	10.30273	7.91246	16.11651	3.58579	3853.00
750	1,677	2.87506	3.22453	9.27072	8.26597	16.47260	3.22970	1697.10
850	1,200	2.92942	3.07918	9.02021	8.58150	16.78405	2.91825	828.42
950	776	2.97772	2.88986	8.60519	8.86682	17.06079	2.64151	438.04
1,050	518	3.02119	2.71433	8.20051	9.12759	17.30985	2.39245	246.86
1,150	357	3.06070	2.55267	7.81296	9.36789	17.53622	2.16608	146.58
1,250	218	3.09691	2.33846	7.24200	9.59085	17.74368	1.95862	90.91
1,350	129	3.13033	2.11059	6.60684	9.79897	17.93516	1.76714	58.50
1,450	70	3.16137	1.84510	5.83304	9.99426	18.11301	1.58929	38.84
1,550	31	3.19748	1.49136	4.75793	10.17821	18.27893	1.42337	26.51

Table 32. Cumulative Frequency Tabulation for Data of Industry II (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$1,650	7	3.21748	0.84510	2.71909	10.35218	18.43449	1.26781	18.53
1,750	1	3.24304			10.51731	18.58093	1.12137	13.22
Totals		42.11003	34.56370	101.43000	127.10310			

Table 33. Cumulative Frequency Tabulation for Data of Industry III

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 250	10,139							
350	10,091							
450	9,013	2.65321	3.95487	10.49310	7.03952	10.76349	4.02717	10646.00
550	5,344	2.74036	3.72787	10.21571	7.50957	11.11704	3.67362	4716.60

Table 33. Cumulative Frequency Tabulation for Data of Industry III (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$ 650	1,906	2.81291	3.28012	9.22668	7.91246	11.41136	3.37930	2395.00
750	1,479	2.87506	3.16997	9.11385	8.26597	11.66349	3.12717	1340.20
850	1,078	2.92942	3.03262	8.88382	8.58150	11.88401	2.90665	806.59
950	783	2.97772	2.89376	8.61681	8.86682	12.07995	2.71071	513.70
1,050	499	3.02119	2.69810	8.15147	9.12759	12.25630	2.53436	342.27
1,150	329	3.06070	2.51720	7.70439	9.36789	12.41659	2.37407	236.63
1,250	224	3.09691	2.35025	7.27851	9.59085	12.56348	2.22718	168.73
1,350	135	3.13033	2.13033	6.66864	9.79897	12.69906	2.09160	123.48
1,450	86	3.16137	1.93450	6.11567	9.99426	12.82498	1.96568	92.402
1,550	56	3.19033	1.74819	5.57730	10.17821	12.94247	1.84819	70.50
1,650	36	3.21748	1.55630	5.00736	10.35218	13.05261	1.73805	54.71
1,750	25	3.24304	1.39794	4.53358	10.51731	13.15630	1.63436	43.09
1,850	18	3.26717	1.25527	4.10118	10.67440	13.25419	1.53647	34.393
1,950	15	3.29003	1.17609	3.86937	10.82430	13.34693	1.44373	27.780

Table 33. Cumulative Frequency Tabulation for Data of Industry III (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$2,250	13	3.35218	1.11394	3.73413	11.23711	13.59906	1.19160	15.546
2,750	7	3.43933	0.84510	2.90658	11.82899	13.95261	0.83805	6.890
3,250	5	3.51188	0.69897	2.45470	12.33330	14.24692	0.54374	3.497
3,750	4	3.57403	0.60206	2.15178	12.79369	14.49905	0.29161	1.957
4,250	2							
4,750	2							
5,250	2							
Totals		62.54465	42.08345	126.80463	196.77489			

Table 34. Cumulative Frequency Tabulation for Data of Industry IV

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
\$ 250	18,139							
350	17,352							
450	16,491	2.65321	4.21724	11.18922	7.03952	14.90313	4.80661	64064.00
550	11,569	2.74036	4.06329	11.13488	7.50957	15.39266	4.31708	20753.00
650	6,802	2.81291	3.83264	10.78087	7.91246	15.80017	3.90957	8120.30
750	4,389	2.87506	3.64237	10.47203	8.26597	16.14927	3.56047	3634.70
850	3,082	2.92942	3.48883	10.22025	8.58150	16.45461	3.25513	1799.40
950	1,845	2.97772	3.26600	9.72523	8.86682	16.72591	2.98383	963.46
1,050	1,189	3.02119	3.07518	9.29070	9.12759	16.97008	2.73966	549.11
1,150	723	3.06070	2.85914	8.75097	9.36789	17.19201	2.51773	349.41
1,250	474	3.09691	2.67578	8.28665	9.59085	17.39541	2.31433	206.22
1,350	272	3.13033	2.43457	7.62101	9.79897	17.58313	2.12661	133.85
1,450	159	3.16137	2.20140	6.95944	9.99426	17.75748	1.95226	89.59
1,550	107	3.19033	2.02938	6.47439	10.17821	17.92015	1.78959	61.60

Table 34. Cumulative Frequency Tabulation for Data of Industry IV (Continued)

Income Class Mark (x)	Cumulative Number of Persons (y) Actual	log x	log y Actual	(log x) (log y)	$(\log x)^2$	(v) (log x)	Computed log y	Computed y Cumulative
$1,650	47	3.21748	1.67210	5.37995	10.35218	18.07265	1.63709	43.36
1,750	12	3.24304	1.07918	3.49982	10.51731	18.21622	1.49352	31.16
1,850	12	3.26717	1.07918	3.52586	10.67440	18.35176	1.35798	22.80
1,950	6	3.29003	0.77815	2.56014	10.82430	18.48016	1.22958	16.966
2,250	3	3.35218	0.47712	1.59939	11.23711	18.82926	0.88048	7.594
2,750	1							
3,250	1							
Totals		52.01941	42.87155	127.47080	159.83891			

BIBLIOGRAPHY

1. Harold T. Davis, The Theory of Econometrics, The Principia Press, Inc., Bloomington, Indiana, 1941, pp. 19, 23, and 26.

2. J. Aitchison and J. A. C. Brown, "On Criteria for Descriptions of Income Distribution," Metroeconomica, Vol. 6, December, 1954, pp. 88-107.

3. R. Gibrat, Les inegalites economiqués, Recueil Sirey, Paris, 1931.

4. H. S. Steyn, "A Model for the Distribution of Incomes," The South African Journal of Economics, Vol. 27, June, 1959, pp. 149-156.

5. A. N. Kolmogoroff, C. R. Acad. Sci., U. R. S. S., Vol. 31, No. 2, 1941.

6. P. R. Halmos, Annals of Mathematical Statistics, Vol. 15, 1944.

7. B. Epstein, Journal of the Franklin Institute, Vol. 244, 1947.

8. D. G. Champernowne, "The Graduation of Income Distributions," Econometrica, Vol. 20, No. 4, 1952, pp. 591-615.

9. E. C. Rhodes, Economica, New Series, Vol. 11, 41, 1944.

10. M. Castellani, Ann. Math. Statist., 21, 1950.

11. Vilfredo Pareto, Cours d'economie politique, 2 Vols., Lausanne, 1896-97.

12. C. Snyder, "The Pareto Curve and its Significance for Our Time," Report of the Research Conference of the Cowles Commission, 1937, pp. 60-61.

13. Miyoji Hayakawa, "The Application of Pareto's Law of Income to Japanese Data," Econometrica, Vol. 19, April, 1951, pp. 174-183.

14. G. F. Shirras, "The Pareto Law and the Distribution of Income," Economic Journal, Vol. 45, 1935, pp. 663-681.

15. A. C. Bose and S. B. Roy, "Application of the Pareto Formula to the Distribution of Personal Income in India," Indian Economic Review, Vol. 3, No. 2, August, 1956, pp. 115-118.

16. R. G. D. Allen, "Changes in the Distribution of Higher Incomes," Economica, New Series, Vol. 24, No. 94, 1957, p. 138.

17. United Nations: Economic Bulletin for Europe, Vol. 8, No. 2, 1956, p. 38.

18. Lee Soltow, "Income Equality in a Factory Payroll," Southern Economic Journal, Vol. 25, January, 1959, pp. 343-348.

19. Kenneth E. Foster, Gerald F. Wajda, and Theodore R. Lawson, "Global Plan for Salary Administration," Harvard Business Review, Vol. 39, No. 5, September-October, 1961, pp. 62-66.

20. National Industrial Conference Board: "Compensation of Top Executives," Studies in Personnel Policy, No. 173, 1959.

21. Plant and Product Directory (The 500 Largest U. S. Industrial Corporations), By the Market Research Department of FORTUNE, 1961.

22. Theodore R. Running, Empirical Formulas, John Wiley and Sons, Inc., New York, 1917.

23. John H. Perry, Chemical Business Handbook, McGraw Hill Book Co., Inc., New York, 1954, pp. 20-55 through 20-77.

24. Yu. V. Linnik, Method of Least Squares and Principles of the Theory of Observations, Pergamon Press, New York, 1961.

25. H. T. Davis and W. F. C. Nelson, Elements of Statistics, The Principia Press, Inc., Bloomington, Indiana, 1937, pp. 215-252.